A MARKOV RANDOM FIELD MODEL-BASED APPROACH TO IMAGE INTERPRETATION†

J. W. MODESTINO AND J. ZHANG‡
ELECTRICAL, COMPUTER AND SYSTEMS ENGINEERING DEPARTMENT
RENSSELAER POLYTECHNIC INSTITUTE
TROY, NEW YORK, 12180

ABSTRACT

In this paper, a Markov random field (MRF) model-based approach to automated image interpretation is described and demonstrated as a region-based scheme. In this approach, an image is first segmented into a collection of disjoint regions which form the nodes of an adjacency graph. Image interpretation is then achieved through assigning object labels, or interpretations, to the segmented regions, or nodes, using domain knowledge, extracted feature measurements and spatial relationships between the various regions. The interpretation labels are modeled as a MRF on the corresponding adjacency graph and the image interpretation problem is formulated as a maximum a posteriori (MAP) estimation rule. Simulated annealing is used to find the best realization, or optimal MAP interpretation. Through the MRF model, this approach also provides a systematic method for organizing and representing domain knowledge through the clique functions of the pdf of the underlying MRF. Results of image interpretation experiments performed on synthetic and real-world images using this approach are described and appear promising.

I. Introduction

Image interpretation is the process of understanding the meaning of an image through identifying significant objects in the image and analyzing their spatial relationships. The need for image interpretation can be found in many diverse fields of science and engineering, such as remote-sensing, or aerial/satellite photointerpretation, for geological survey and military air reconnaissance [1]- [3], [8], biomedical science and particle physics [4]-[5], and robot vision systems [6]-[7].

Most of the existing image interpretation techniques involve two major operations, *low-level* and *high-level* processing. In low-level processing, the representation of an image is transformed, through image processing operations, such as edge detection and region segmentation, from a *numerical* representation, as an array of pixel intensities, to a *symbolic* representation, as a set of spatially related *image primitives*, such as edges and regions. Various features are then extracted from the primitives. These features may include: the lengths of significant edges, average intensities of regions, shape and/or texture descriptors, etc. Also extracted would be the spatial relationship between the image primitives. In high-level processing, image domain knowledge is used to assign object labels, or interpretations, to the primitives and construct a description as to "what is present in the image". In the rest of this paper, we often refer to the object labels as interpretations and the overall interpretations for *all* the primitives in the image as the interpretation of the image.

†This work was supported in part by USAF RADC under Contract No. F30620-82-K-0151.

‡Presently with the Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

Due to the very complex nature of the image interpretation problem, most of the recent techniques have adopted the *knowledge-based*, or *expert system*, approach. In this approach, domain knowledge, and especially spatial constraints, are used in high-level (some also in low-level) processing. Hence, an ambiguous object may be recognized as the result of successful recognition of its neighboring objects. Even more fundamentally, an object can be recognized from combining the feature information from several spatially-related image primitives. Finally, low- level processing errors may be corrected, or at least mitigated, through feedback from high-level processing to low-level image processing.

The early work in knowledge-based image interpretation has been summarized in Nagao and Matsuyama [11], Binford [12] and Ohta [13]. Recently, a number of more sophisticated experimental systems have been constructed for different application domains, such as high-altitude aerial photographs [11], [14]-[15], [16]-[18]; airport scenes [19], [20]-[21] and outdoor scenes [13],[22]-[24]. Many of these systems are still undergoing continuous improvements through architecture modification and domain extension. New ideas and systems are constantly emerging, as can be seen in recent PRCV and SPIE conferences and workshops, and several pertinent technical reports [25]-[26],[21]. While success has been demonstrated to various degrees in these systems, developing a *general, domain-independent* and *systematic* method for constructing knowledge-based image interpretation systems is still an open problem [22].

In this paper, we describe a general, domain-independent, stochastic model- based approach to the image interpretation problem. In this approach, the interpretation labels to be assigned to the primitives of an image are modeled as a Markov random field (MRF) defined on the spatial adjacency graph formed by the primitives, where the randomness is used to model the uncertainty in the assignment of the labels. As a result, the domain knowledge, whatever it may be, can be systematically represented in terms of the clique functions associated with the underlying Gibbs probability distribution function (pdf) describing the MRF. Under the MRF modeling assumption, image interpretation is then formulated as the optimization problem of maximizing the a posteriori probability of interpretation given domain knowledge and feature measurements. Then, simulated annealing is used to find the optimal set of interpretation labels. In this paper, we present a special region-based version of this approach. That is, the primitives are segmented regions; and for the sake of simplicity, we do not include feedback from high-level to low-level processing. However, research is currently ongoing to include use of linear edge segments as primitives as well as high-to-low level feedback [27]-[28]. This will also be discussed in subsequent sections.

After describing the MRF model-based formulation in the next section, we will show how domain knowledge can be organized into clique functions associated with the MRF model in Section III. The implementation of the optimization (interpretation) through simulated annealing is described in Section IV. We will present and discuss results of image interpretation experiments performed on synthetic and real-world images in Section V. Finally, a summary and directions for future research are provided in Section VI.

II. The MRF Model-Based Approach

The MRF model has recently attracted much attention in the image processing and computer vision community. The main advantage of this model is that it provides a general and natural model for the interaction between spatially related random variables and there is a relatively efficient optimization algorithm, simulated annealing, that can be used to find the *globally* optimal realization. Up to now, the success of MRF models has been demonstrated mostly in low-level image processing applications, such as region segmentation [29]-[36] and edge detection [37], where they are defined on the two dimensional (2-D) image lattices. However, as demonstrated by Kinderman and Snell [38], the MRF can be defined, in general, on graphs for which the 2-D lattice is a special case. In what follows, we will briefly review the concepts associated with the MRF defined on graphs and show how this can be applied to the image interpretation problem. More comprehensive treatments on MRF's can be found in [38]-[39].

A.) The MRF Model on Graphs:

Let $\mathbf{G} = \{\mathbf{R}, \mathbf{E}\}$ be a graph, where

$$\mathbf{R} = \{R_1, R_2, \ldots, R_N\} \tag{1}$$

is the set nodes represented by R_i, $i = 1, 2, \ldots, N$; $\mathbf{E}$ is the set of edges connecting them. Suppose that there exists a *neighborhood system* on $\mathbf{G}$, denoted by

$$\mathbf{n} = \{n(R_1), n(R_2), \ldots, n(R_N)\}, \tag{2}$$

where $n(R_i)$, $i = 1, 2, \ldots, N$, is the set of all the nodes in $\mathbf{R}$ that are neighbors of R_i, such that

i.) $R_i \notin n(R_i)$, and

ii.) if $R_j \in n(R_i)$ then $R_i \in n(R_j)$.

Let

$$\mathbf{I} = \{I_1, I_2, \ldots, I_N\} \tag{3}$$

be a family of random variables defined on $\mathbf{R}$. Then, $\mathbf{I}$ is called a *random field*, where I_i is the random variable associated with R_i. Notice that the random variables I_i's here can be numerical as well as symbolic, e.g., interpretation labels. We say $\mathbf{I}$ is a MRF on $\mathbf{G}$ with respect to the neighborhood system $\mathbf{n}$ if and only if

i.) $P[\mathbf{I}] > 0$, for all realizations of $\mathbf{I}$;

ii.) $Pr[I_i | I_j, \text{all } R_j \neq R_i] = P[I_i | I_j, R_j \in n(R_i)]$,

where $P[\cdot]$ and $P[\cdot|\cdot]$ are the joint and conditional pdf's, respectively. Intuitively, the MRF is a random field with the property that the statistics at a particular node depends mainly on that of its neighbors.

An important feature of the MRF model defined above is that its joint pdf has a general functional form, known as the *Gibbs distribution*, which is defined based on the concept of *cliques* [38]-[39]. Here, a clique associated with the graph $\mathbf{G}$, denoted by c, is a subset of $\mathbf{R}$ such that it contains either a single node or several nodes that are *all* neighbors of each other. If we denote the collection of all the cliques of $\mathbf{G}$ with respect to the neighborhood system $\mathbf{n}$ as $\mathbf{C}(\mathbf{G}, \mathbf{n})$, the general functional form of the pdf[1] of the MRF can be expressed as the following Gibbs distribution:

$$P[\mathbf{I}] = Z^{-1} exp[-U(\mathbf{I})], \tag{4a}$$

where

$$U(\mathbf{I}) = \sum_{c \in \mathbf{C}(\mathbf{G}, \mathbf{n})} V_c(\mathbf{I}) \tag{4b}$$

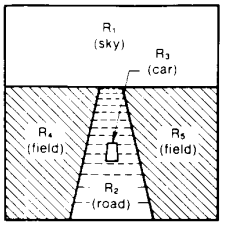
is called the *Gibbs' energy function* and $V_c(\mathbf{I})$'s are called clique functions defined on the corresponding cliques $c \in \mathbf{C}(\mathbf{G}, \mathbf{n})$. Finally,

$$Z = \sum_{\text{all } \mathbf{I}'} exp\Big[-U(\mathbf{I}')\Big] \tag{4c}$$

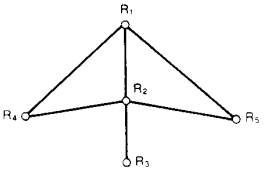
is the normalization factor to make (4a) a valid pdf. Notice that the MRF pdf above is quite rich in that the clique functions can be arbitrary as long as they depend only on the nodes in the corresponding cliques. Due to this unique structure, in which the global and local properties are related through cliques, the MRF model-based approach to image interpretation provides potential advantages in knowledge representation, learning and optimization, as will be discussed in more detail later. More importantly, this method provides a useful mathematical framework for the study of image interpretation procedures.

B.) The MRF Model-based Formulation:

As described in Section I, for the time being we restrict the image interpretation problem to that of labeling *segmented regions*. Suppose for a given image, there are N disjoint regions after segmentation,[2] denoted by $\mathbf{R} = \{R_1, R_2, \ldots, R_N\}$. Then $\mathbf{R}$ can be represented by a set of nodes in a connected graph, called the *adjacency graph*, denoted by $\mathbf{G} = \{\mathbf{R}, \mathbf{E}\}$; where the edge set $\mathbf{E}$ is such that a node R_i is connected to another node R_j if and only if the corresponding regions are spatially adjacent. A neighborhood system, denoted by $\mathbf{n}$, can also be defined on the adjacency graph. For simplicity, in what follows we define the neighbors of a node to be the nodes that are connected to it directly by an edge of $\mathbf{G}$, i.e., only spatial adjacent regions are neighbors. Now, given the neighborhood system, we can also find the cliques for the adjacency graph. As an illustration, we have shown in Fig. 1 the adjacency graph and all its cliques for a particular synthetic conceptual image. This image is intended to represent a car on a road between two fields with the sky as a background. In forming the adjacency graph, we assume perfect segmentation of the image objects.



a. The synthetic conceptual image



b. the adjacency graph of the conceptual image

Node or Region	Associated Cliques
R_1	$\{R_1\}$, $\{R_1, R_2\}$, $\{R_1, R_4\}$, $\{R_1, R_5\}$ $\{R_1, R_2, R_4\}$, $\{R_1, R_2, R_5\}$
R_2	$\{R_2\}$, $\{R_2, R_3\}$, $\{R_2, R_4\}$, $\{R_2, R_5\}$
R_3	$\{R_3\}$
R_4	$\{R_4\}$

c. cliques of the adjacency graph

Figure 1 A Synthetic Conceptual Image for Image Interpretation

As described in Section I, image interpretation is the process of assigning object labels to the segmented regions according to domain knowledge and feature measurement information (or *measurements*, in short) made on these regions. From the above graphical formulation, the interpretation of the image can be represented as a vector $\mathbf{I}(\mathbf{R}) = \{I_1, I_2, \ldots, I_N\}$, defined on the adjacency graph $\mathbf{G}$, where we use $\mathbf{I}(\mathbf{R})$ to emphasize the relationship between interpretation and the symbolic representation in terms of segmented regions. Here, I_i, $i = 1, 2, \ldots, N$, is the interpretation label for node R_i; while $I_i \in L$

[1] Actually, this is a probability mass function (pmf) due to the discrete nature of $\mathbf{I}$ although we will not make this distinction in what follows and continue to use the term pdf.

[2] Clearly, the number N of segmented regions is a random variable depending upon the image as well as the segmentation procedure.

and $L = \{L_1, L_2, \ldots, L_M\}$ is the set of all the interpretation labels. In addition, we consider I_i's as symbolic random variables to account for the uncertainty in assigning object labels to segmented regions due to, e.g., image noise and segmentation errors. Hence, $\mathbf{I}(\mathbf{R})$ is a random field. Let's denote the domain knowledge as $\mathbf{K}$ and all the measurements made on the segmented regions as $\mathbf{X}(\mathbf{R})$. Now, we can define image interpretation as the following *optimization problem*: for a given $\mathbf{R}$, find $\mathbf{I}_0(\mathbf{R})$, such that

$$\mathbf{I}_0(\mathbf{R}) = \arg \max_{\mathbf{I} \in \{L\}^N} P[\mathbf{I}(\mathbf{R}) \,|\, \mathbf{K}, \mathbf{X}(\mathbf{R})], \tag{5}$$

where $P[\cdot|\cdot,\cdot]$ is the a posteriori pdf of the interpretation given the domain knowledge and measurements, while $\{L\}^N$ is the set of all possible interpretation vectors of length N. The formulation of (5) is also known as the maximum a posteriori (MAP) formulation.

Two problems must be solved in applying the above MAP approach to image interpretation. Specifically, we need an explicit expression for the conditional pdf in (5) and an optimization method to avoid the computationally explosive nature of exhaustive combinatorial search. Feldman and Yakimovski [40], and Faugeras and Price [14]-[15] have considered similar formulations to that of (5) and proposed heuristic expressions for the a posteriori pdf using the marginal pdf's of single and joint pdf's of pairs of interpretation labels. They have also used different relaxation schemes to find local optimal solutions, some of which have also been studied in [54]-[57]. On the other hand, the MRF model discussed in A.) appears to provide a natural solution to the above two problems. More specifically, assume that $\mathbf{I}(\mathbf{R})$ forms a MRF. Then, the pdf appearing in (5) is the Gibbs distribution

$$P[\mathbf{I}(\mathbf{R}) \,|\, \mathbf{K}, \mathbf{X}(\mathbf{R})] = Z^{-1} exp[-U(\mathbf{I}(\mathbf{R}) \,;\, \mathbf{K}, \mathbf{X}(\mathbf{R}))], \tag{6a}$$

with energy function

$$U(\mathbf{I}(\mathbf{R}) \,;\, \mathbf{K}, \mathbf{X}(\mathbf{R})) = \sum_{c \in \mathbf{C}(\mathbf{G}, \mathbf{n})} V_c(\mathbf{I}(\mathbf{R}) \,;\, \mathbf{K}, \mathbf{X}(\mathbf{R})), \tag{6b}$$

where the $V_c(\cdot\,;\cdot,\cdot)$'s are the clique functions. Indeed, as will be seen in the subsequent sections, through imposing a neighborhood system and the Markov property of ii.), the MRF model-based formulation provides a general and systematic approach for knowledge representation and knowledge acquisition through appropriate construction of the clique functions. For the optimization strategy, the simulated annealing procedure can be used to find the globally optimal interpretation for the image. In addition, the approach of [40],[14],[15] can be shown to be special cases with certain neighborhood structures and clique functions. Finally, when used in the context of image interpretation, the MRF model suggest that the interpretation for a particular region given those of all other regions, depends only on the interpretations of its neighboring regions. This is often a reasonable assumption in practical applications. For example, the identification of a region as a car might depend on whether its neighboring regions are a road but has little to do with the identity of the regions spatially far removed from it. In the rest of the paper, we will model the interpretation vector $\mathbf{I}(\mathbf{R})$ as a MRF. Since simulated annealing is a relatively well-defined procedure, we will concentrate on the knowledge engineering aspects of the image interpretation problem; that is, knowledge representation and learning through constructing the pdf of the MRF.

Finally, we should point out that, although the MRF model-based approach is presented here in the form of a region-based approach, it can be extended to include other primitives and to model the situation where interaction between high-level and low-level processing (e.g., feedback) is used. When other primitives, such as linear edge segments, are introduced they can be considered as nodes of a generalized adjacency graph; they can also be considered as features associated with different regions rather than primitives themselves. To model the high-level and low-level interaction during interpretation, the adjacency graph can be considered as a *dynamic* graph which changes with time; subsequently, the MRF become also a dynamic model. Currently, these problems are under active investigation.

III. Clique Function Design

In the MRF model-based formulation of the preceding section, it is clear that the optimal interpretation, $\mathbf{I}_0(\mathbf{R})$, should be the one that minimizes the energy function, or has the minimum energy. For a given image, the optimal interpretation depends on how the energy function is defined. In general, we would like the optimal interpretation obtained under the MRF assumption to be the one that is most *consistent* with the measurements and domain knowledge. For example, in aerial photointerpretation, suppose we know that a car has small area and would usually be on a road. An interpretation with a car having large area or in the sky should obviously be considered *not* optimal. This type of consistency requirement can be achieved by properly selecting the energy functional or, rather, the corresponding clique functions. Without loss of generality, we assume that all the clique functions are non- negative. Then, a general principle for the selection of a clique function, which maintains the consistency between the interpretation, knowledge, and measurements, is the following.

Design Rule:

If the interpretation of the regions (or region for a *singleton* clique) in a clique tends to be consistent with the measurements and domain knowledge, the clique function decreases, resulting in a decrease in the energy function; otherwise, the clique function increases, resulting in a corresponding increase in the energy function.

In this way, an interpretation for the image that is most consistent with the measurements and domain knowledge will have the minimum energy, or achieve the optimum. Based on this principle, we now propose a general approach to defining clique functions from domain knowledge. We first consider the clique functions for single-node cliques, and then extend the result to the case of multiple-cliques.

A.) Clique Functions for Single-Node Cliques:

Let c be an arbitrary single-node clique with one node, R. Let the corresponding clique function be denoted by $V_c(I(R) \,;\, \mathbf{K}, \mathbf{X}(R))$, it depends only upon the single node R, its interpretation $I = I(R)$, and the measurements $\mathbf{X}(R)$ on the corresponding segmented region R, as well as the domain knowledge represented by $\mathbf{K}$. Suppose that $\mathbf{X}(R)$ has m components, $X_1(R), X_2(R), \ldots, X_m(R)$, representing measurement values of m well- defined *features* of R, e.g., *average gray level, area, standard deviation of gray-levels*, etc. Assuming the components of $\mathbf{X}(R)$ are independent, we can define a clique function for clique c as

$$\begin{aligned} &V_c(I(R) \,;\, \mathbf{K}, \mathbf{X}(R)) \\ &= \sum_{i=1}^{m} p_c^{(i)}(I(R), \mathbf{K}) B_c^{(i)}(I(R); \mathbf{K}, X_i(R)), \end{aligned} \tag{7}$$

where $B_c^{(i)}(\cdot\,;\cdot,\cdot)$, $i = 1, 2, \ldots, m$, are called *basis functions* for the corresponding clique function. These quantities are functions of the i'th feature measurement, $X_i(R)$, parameterized by the interpretation $I(R)$ and, of course, depend upon the domain knowledge, $\mathbf{K}$. The $p_c^{(i)}(I, \mathbf{K})$'s, are a set of non negative numbers, or *weights*, associated with the basis functions, i.e.,

$$p_c^{(i)}(I, \mathbf{K}) \geq 0 \;;\; i = 1, 2, \ldots, m, \tag{8a}$$

and can be conveniently normalized so that

$$\sum_{i=1}^{m} p_c^{(i)}(I, \mathbf{K}) = 1. \tag{8b}$$

Here, $p_c^{(i)}(I, \mathbf{K})$ depends not only on i but also on the interpretation $I(R)$ as well as $\mathbf{K}$.

Now the problem of designing clique functions becomes that of designing the basis functions of the features and determining their weights. We first consider the design of the basis functions. Without loss of generality, we assume that all the basis functions are non-negative. Then, the consistency principle for designing clique functions (in the previous Design Rule) applies to the design of the basis functions. Here, it is sufficient to consider the design of a particular basis function for a single-node clique c, denoted by $B_c\big(I(R); \mathbf{K}, X(R)\big)$, where, for notational simplicity, the index i has been dropped. According to the consistency requirement between interpretation, measurements and domain knowledge, we want the basis function to be small when $I(R)$, $X(R)$ are *consistent* according to $\mathbf{K}$; otherwise, it should be large. One way to achieve this is to take a probabilistic approach. In particular, we consider the a posteriori pdf $P_c[I(R) \mid \mathbf{K}, X(R)]$. This is the probability that, based on the domain knowledge, $\mathbf{K}$, and the measurement, $X(R)$, the interpretation of the node R should be $I(R)$. By definition, the probability $P_c[I(R) \mid \mathbf{K}, X(R)]$ is such that for $I(R)$ consistent with the measurements and domain knowledge, it is large; otherwise it is small. Hence, a non-increasing function of this pdf can be used as a basis function. For example, the logarithm of the pdf has been suggested [38] as a reasonable basis function for general MRF's. In this case we can define

$$B_c\big(I(R); \mathbf{K}, X(R)\big) = -\alpha_c log P_c[I(R) \mid \mathbf{K}, X(R)], \tag{9}$$

where α_c is a positive weighting constant and $-log(\cdot)$ is a monotonically decreasing function. Another way of selecting the basis function is to use

$$B_c\big(I(R); \mathbf{K}, X(R)]\big) = \alpha_c\big(1 - \beta_c P_c[I(R) \mid \mathbf{K}, X(R)]\big), \tag{10}$$

where α_c and β_c are positive constants, and $\beta_c P_c[I(R) \mid \mathbf{K}, X(R)] < 1$. Usually, we want the normalization constants α_c and β_c to be such that $0 \leq B_c\big(I(R); \mathbf{K}, X(R)\big) \leq 1$.

To find the pdf $P_c[I(R) \mid \mathbf{K}, X(R)]$, Bayes' conditional pdf formula can be used. That is,

$$\begin{aligned} P_c[I(R) \mid \mathbf{K}, X(R)] = & P_c[X(R) \mid \mathbf{K}, I(R)] P_c[\mathbf{K}, I(R)] \cdot \\ & \cdot P^{-1}[\mathbf{K}, X(R)], \end{aligned} \tag{11}$$

where the terms on the right hand side can be obtained through proper modeling assumptions [46].

In Fig. 2, we have shown several types of basis functions obtained using the probabilistic approach described above, while leaving the details of the derivation and more examples in our recent work [46],[51],[58]. Notice that these functions are all "window-like" functions. When domain knowledge about a feature can be expressed in terms of a nominal value or an interval of values, the specification of the corresponding basis function can be greatly simplified to that of merely constructing one of these window functions. The piecewise linear basis function is particularly interesting in that it is very easy to compute and, it is relatively robust against measurement errors or image noise [46]. Hence, we give it a special notation, $g(x; a_1, a_2, b_1, b_2)$; where x is the variable and a_1, a_2, b_1, b_2 are the four "corner points", with $a_1 \leq a_2 \leq b_1 \leq b_2$. Similar functions have been used in [22] for a rule-based image interpretation system and in the applications of fuzzy set theory [42].

B.) Clique Functions for Multiple-Node Cliques:

Here, we still design clique functions through designing a set of basis functions, as indicated in expression (7). However, the designing of the basis function is slightly more complicated here in that we may have two types of basis functions. The first type is the basis function for feature measurements, as in the case for single-node cliques. The features in this case could be quantities such as *mutual boundary length, contrast*, etc. Basis functions for these feature measurements can be designed in the same way as that in *A.)* using the window functions of Fig. 2. The second type of basis functions are those for spatial constraints. The constraints in this case could be statements such as "a car should be on (neighboring to) the road", "a car should never be in the sky", etc. In this case, we can still use the probabilistic approach in the spirit of (9)-(10). For example, consider an arbitrary clique c with multiple nodes denoted by $\mathbf{R}_c$ and interpretations $\mathbf{I}_c(\mathbf{R}_c)$. Let $P_c[\mathbf{I}_c(\mathbf{R}_c) \mid \mathbf{K}]$ be the probability that the combination of interpretations $I_c(\mathbf{R}_c)$ is valid according to domain knowledge. For example we might have

$$\begin{aligned} P_c[\mathbf{I}_c(\mathbf{R}_c) \mid \mathbf{K}] = 1; & \quad \text{if } \mathbf{I}_c \text{ is a valid combination,} \\ = 0; & \quad \text{if } \mathbf{I}_c \text{ is not a valid combination.} \end{aligned}$$

Similar to (9)-(10), we can define the basis function as

$$B_c(\mathbf{I}_c(\mathbf{R}); \mathbf{K}) = \alpha_c(1 - P_c[\mathbf{I}_c(\mathbf{R}) \mid \mathbf{K}]). \tag{14}$$

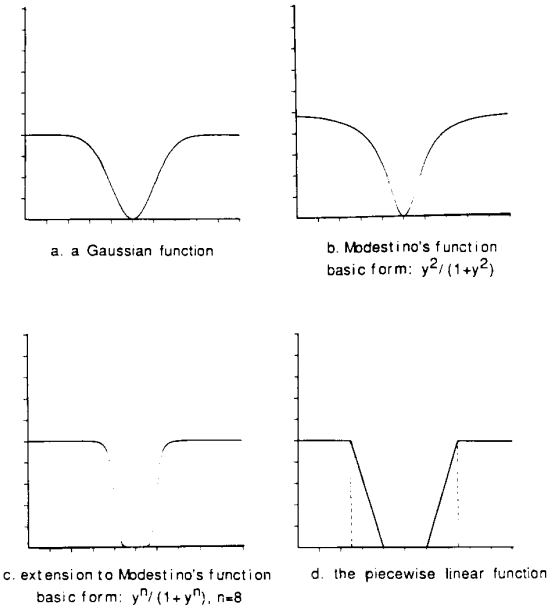


Figure 2 Example of Different Basis Functions

C.) The Selection of the Weights for the Basis Functions:

The weights of the basis functions in (7)-(8) control the contributions of the individual basis functions to the value of a clique function. For simplicity, we may make them all equal. In our current experiments we start with this simple scheme and then, if a feature is too unreliable for a particular object type, we will reduce the corresponding weight. In addition, adjustments are also made by trial-and-error through examining interpretation results on representative training images. A more sophisticated approach, currently under investigation, is to select a weight based on how powerful the corresponding feature is for object recognition and discrimination according to well-defined criteria [43].

D.) Remarks:

Through the design of clique functions, we have a systematic approach for representing spatial knowledge; that is, for organizing the domain knowledge into a set of well-defined clique functions. This approach also provides guidelines as to what kind of knowledge one would want for the purpose of image interpretation; basically, knowledge concerning objects spatially related as members of different type of cliques. It seems that many of the "rules" in the previous expert systems mentioned in Section I can be transformed into clique functions, where the "condition" parts correspond to evaluating clique functions and the "action" parts correspond to assigning labels. In fact, it is shown by Smolensky [60], that in many cases, the set of rules in an expert system can be represented by a properly defined Gibbs distribution.

IV. Implementation

Under the MRF modeling assumption, image interpretation becomes the problem of finding the optimal interpretation vector that minimizes the energy function. A simple exhaustive search procedure will result in an exponential complexity of $O(M^N)$, where M is the number of labels and N is the number of nodes in the adjacency graph. Hence, we make use of the simulated annealing algorithm, a stochastic iterative optimization procedure, that will find the global maximum of the pdf of the MRF, or the minimum of the energy function, without excessive computation [29],[48].

The simulated annealing algorithm has been widely used in various applications involving combinatorial optimization, such as VLSI layout [44], channel coding [45], image segmentation [46]-[47]. For convenience, we rewrite this algorithm here in the context of a minimization problem. Let the function to be minimized be $E(\mathbf{x})$, where $\mathbf{x}$ is the indepdendent variable. This algorithm can be loosely described as follows:

Simulated Annealing

1) Select an initial "temperature" parameter T_0 and randomly choose an initial variable $\mathbf{x}_0$. Iteration begins.
2) At step k, perturb $\mathbf{x}_k$ by $\hat{\mathbf{x}}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}$ and compute $\Delta E = E(\hat{\mathbf{x}}_{k+1}) - E(\mathbf{x}_k)$.
3) If $\Delta E < 0$ accept the change; that is

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}.$$

If $\Delta E > 0$, accept the change only with probability $p = e^{-\Delta E/T}$.
4) If there is a considerable drop in energy, or enough iterations, lower the temperature.
5) If the energy becomes stable and the temperature is very low, stop; otherwise go back to (2).

When applied to image interpretation, the function $f(\cdot)$ is the energy function, and the variables are the set of interpretation labels. An iteration is defined as *one* visit to *all* the nodes and perturbation is achieved through generating a new label for a node when it is visited from an uniform distribution of all object labels. Finally, the temperature is lowered after each iteration according to $T_{l+1} = \alpha T_l$ where $0.5 < \alpha < 1$, In particular, we have selected $T_0 = 1$ and $\alpha = 0.92$ for all our experiments. More details of the implementation can be found from [46],[58].

V. Experimental Results

To test the efficacy of the MRF model-based approach, we have performed some preliminary experiments on synthetic and real-world images [46], [51], [52], [58]. In what follows, we will summarize these experimental results and discuss their implications. Due to the space constraint of the paper, we will only present a few representative results while more detailed accounts can be found in [46], [51], [52], [58].

A.) Interpretation of Synthetic Images:

The efficacy of the MRF model-based approach depends on the validity of the MRF assumption, on the quality of the segmentation, on how powerful the features and spatial constraints in the knowledge are (as far as object recognition is concerned), and finally, on how effective the simulated annealing is (convergence of interpretations). Before applying the MRF approach to real-world image interpretation, then, it makes sense to test whether this approach would work at all under the somewhat ideal situation in which we have acceptable segmentation and relatively strong features and spatial constraints in the domain knowledge. If the MRF model-based approach works well here, then it is reasonable to predict that it will be effective for real-world image interpretation provided we can produce good segmentation (or are able to deal with a poor one) and find strong features and spatial constraints. We can create such an ideal situation by generating synthetic images and studying the performance of the MRF model-based approach in interpreting them, as illustrated by the typical results shown in Fig. 3.

The synthetic images used in this experiment is a variation from the conceptual image of Fig. 1 which contains such objects as sky, road, field, and car, all appearing as regions of *constant* gray-levels. As for the variation, additive white Gaussian noise (3dB signal-to-noise ratio (SNR)) is added to the image and an unknown object which has been deliberately made to have a similar area and average gray-level to that of the car is planted in the image in order to simulate the effects of segmentation error, measurement error and unknown objects which have no descriptions in the knowledge base. The assumed domain knowledge for this set of images is stated in Table 1, the precise definitions of the object features are in Table 2, and the clique function representation of the domain knowledge is shown in Table 3 in terms of the piecewise-linear basis functions and their weights.

Starting from a random interpretation, the simulated annealing has converged to the correct interpretation, i.e., all the "objects" have been correctly labeled, within 25 iterations for Fig. 3, where the object labels are represented by different constant gray tones. The result here, along with many others in [46], [58], indicates that with relatively strong features and spatial constraints, the MRF model-based approach using simulated annealing is quite effective. In particular, although relatively large measurement errors occurred to the region corresponding to the car due to noise and poor segmentation, and the unknown object has feature measurements very close to those of the regions corresponding to the car, spatial constraints, represented in the clique functions involving two or three nodes, led to correct interpretations.

B.) Interpretation of Real-World Images:

To study its practical applicability, the MRF model-based approach is tested on real-world images, in particular, aerial photographs. The main difference between experiments on real-world images and those on synthetic images is that for real-world images, the knowledge, or the clique functions, has to be obtained from knowledge acquisition, or *learning*. The MRF formulation has provided a general structure as to what kind of knowledge we should look for, i.e., knowledge related to the cliques. A knowledge source that is commonly used, and is also used here, is a set of training images in which the objects and their spatial relationships are identified by human experts.

A typical training image is shown in Fig. 4, which contains mainly the following objects — vegetation region (VEGE), shadow (SHD1), shadow on the ground (SHD2), ground (GRND) and oil tank (OLTK). The knowledge in this case is represented in terms of the piecewise basis functions in Table 4. Notice here, when the types of features and spatial constraints are selected, the problem of learning becomes that of determining the corner points and weights for the basis functions based on the measurements from the identified objects in the training image. Although the corner points for the piecewise-linear basis functions of Table 4 are determined heuristically (see [58]), we have recently developed and implemented an approach to automate the process [59]. Some interpretation results using the knowledge of Table 4 are shown in Figs. 4-5. More specifically, Fig. 4c is the computer interpretation of the training image done to verify the correctness of the acquired knowledge, and Fig. 5c is the computer interpretation of a test image which contains basically the same type of objects as those in the training image. In both cases, most of the objects that *have* descriptions in the knowledge base are correctly identified. In addition, although the segmentation is quite poor for the test image due to the condition of the original image (see Figs. 5a and 5b, in which some oil tanks are very difficult to separate from the background even for a human observer), the interpretation is still reasonably good due to the use of relatively powerful features (e.g., the partial compactness for noisy and partial circles) and spatial constraints [52]. The simulated annealing converged in both case within 50 iterations. Our preliminary results [46], [51], [52], [58] have shown that the MRF model-based approach is quite effective for practical applications provided that proper features and constraints are used in the clique functions.

VI. Summary

In this paper, a Markov random field (MRF) model-based approach to automated image interpretation is described and demonstrated as a region-based scheme. In this approach, an image is first segmented into a collection of disjoint regions which form the nodes of an adjacency graph. Image interpretation is then achieved through assigning object labels, or interpretations, to the segmented regions, or

nodes, using domain knowledge, extracted feature measurements and spatial relationships between the various regions. The interpretation labels are modeled as a MRF on the corresponding adjacency graph and the image interpretation problem is formulated as a maximum a posteriori (MAP) estimation rule. Simulated annealing is used to find the best realization, or optimal MAP interpretation. Through the MRF model, this approach also provides a systematic method for organizing and representing domain knowledge through the clique functions of the pdf of the underlying MRF. In particular, we have proposed a general structure for the clique functions as a weighted sum of basis functions of features and spatial constraints Image interpretation experiments performed on synthetic and real-world images using this approach are described and appear promising.

While the results here are still preliminary, they do suggest a promising direction for future research work. Specifically, future research should include the following three immediate research tasks, namely, finding a general approach to determine the weights and corner points for the piecewise-linear basis functions, incorporate other primitives, such as linear edge segments, and high-level to low-level feedback, such as the split-and- merge of original segmented regions during interpretation, into the MRF model- based approach.

Table 2. Region-Based Features

a.) Features for a Single Region R

1. Area: A = no. of pixels in region R.
2. Average Gray Level: $G = \sum_{(i,j)\in R} x(i,j)/A$,
   where $x(i,j)$'s are gray levels of the pixels in R.
3. Standard Deviation of Gray Levels:
   $S = \left(\sum_{(i,j)\in R}(x(i,j) - G)^2/A\right)^{1/2}$.
4. Compactness (Kanai's): $C = (P^2/4\pi A) - 1$,
   where P is the perimeter of R,
   P = no. of boundary points of R.
5. Partial Compactness (Yu's): C_p = sample stand. dev. of r,
   where r is the random radius of region R.

b.) Features for Two Adjacent Regions R_i and R_j

1. Boundary Length: $B_{i,j} = (P_i + P_j)/2$.
2. Contrast: $C_{i,j} = |G_i - G_j|$.

Table 1 Summary of Assumed Knowledge for the Conceptual Image

a. a single region

	Region Knowledge	
Object Type	Area (No. of pixels)	Average Gray Level
Car	$\leq$ 800	$\doteq$ 150
Sky	$\geq$ 25000	$\doteq$ 200
Road	$\doteq$ 11700	$\doteq$ 100
Field	$\geq$ 13500	$\doteq$ 50

b. two regions

			Mutual Knowledge
Object Type	Boundary Length	Contrast	Spatial Constraints
Sky, Car	0	—	impossible combination
Field, Car	0	—	impossible combination
Sky, Road	$\doteq$ 56	$\doteq$ 100	valid combination
Car, Road	$\doteq$ 120	$\doteq$ 50	valid combination
Sky, Field	$\doteq$ 100	$\doteq$ 150	valid combination
Road, Field	$\doteq$ 180	$\doteq$ 50	valid combination

c. three regions

	High - Order Knowledge
Object Types	Spatial Constraints
Sky, car, field	impossible combination
Sky, car, road	impossible combination
Sky, road, field	valid combination
Road, car, field	impossible combination

Table 3 Linear Basis Functions for the Synthetic Image.

a. cliques of a single node

	Features	
	Area	Average Gray Level
Label	a1, a2, b1, b2, p	a1, a2, b1, b2, p
car	750,790,810,850,0.5	165,175,185,187.0.5
road	11500,11600,11900,12000,0.5	85,95,105,115,0.5
sky	19900,20000,65536,65536,0.5	193,195,205,215,0.5
field	13300,13400,13900,14000,0.5	0,0,55,65,0.5

b. cliques of two nodes

	Features		
	Boundary length	Contrast	Spatial Constraints
Labels	a1, a2, b1, b2, p	a1, a2, b1, b2, p	αc, p
sky, car			1.0, 1.0
field, car			1.0, 1.0
sky, road	0,0,55,65,0.5	85,95,105,115,0.5	0.0, 1.0
car, road	105,115,125,135,0.5	0,0,85,95,0.5	0.0, 1.0
road, field	140,150,1000,1000,0.5	35,45,55,65,0.5	0.0, 1.0
sky, field	85,95,105,115,0.5	140,150,257,257,0.5	0.0, 1.0
all others	—	—	0.5, 1.0

c. three regions

Labels	αc, p
sky, car, field	1.0, 1.0
sky, car, road	1.0, 1.0
sky, road, field	0.0, 1.0
road, car, field	1.0, 1.0
all others	0.5, 1.0

Table 4 Knowledge and Clique Functions for the Aerial Photos.

a. basis function for cliques of a single node

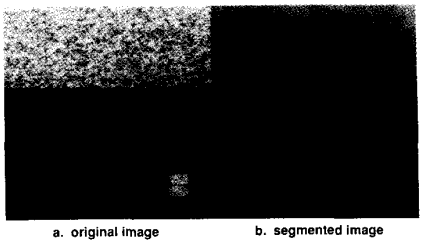
	Features	
	Area	Average Gray Level
Label	a1, a2, b1, b2, p	a1, a2, b1, b2, p
vegetation	11500,12000,13000,13500,0.5	120,123,127,130,0.5
shadow 1	-	135,140,150,155,1.0
shadow 2	0,0,6500,7000,0.5	105,110,140,145,0.5
ground	-	185,190,200,205,1.0
oil tank	4500,4500,64000,64000	200,205,256,256,0.33

a. basis function for cliques of a single node (cont.)

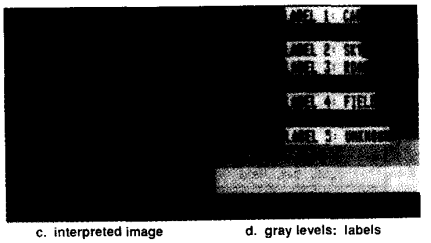
	Features	
	Compactness	Partial Compactness
Label	a1, a2, b1, b2, p	a1, a2, b1, b2, p
vegetation	-	-
shadow 1	-	-
shadow 2	-	-
ground	-	-
oil tank	0,0,0,0,0.95,0.95, 1.0	00,00,100,100,1.0

b. basis function for features and spatial constraints for cliques of two nodes

	Feature	Spatial Constraint
	Contrast	
Lables	a1, a2, b1, b2, p	αc, p
vegetation and shadow 1	10,15,20,30,1.0	0.0, 1.0
shadow1 and shadow 2	0,5,30,35,1.0	0.0, 1.0
shadow 1 and ground	35,40,60,65,1.0	0.0, 1.0
shadow 1 and oil tank	45,50,75,80,1.0	0.0, 1.0
shadow 2 and ground	55,60,90,95,1.0	0.0, 1.0
shadow 2 and oil tank	75,80,105,110,1.0	0.0, 1.0
ground and oil tank	0,5,25,30,1.0	0.0, 1.0
vegetation and oil tank	impossible combination	1.0, 1.0
oil tank and oil tank	impossible combination	1.0, 1.0
all other cases		0.5, 1.0



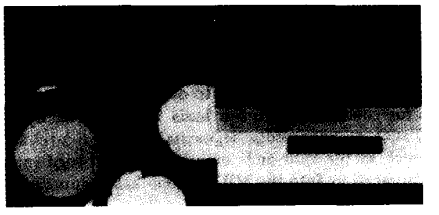
a. original image b. segmented image



c. interpreted image d. gray levels: labels

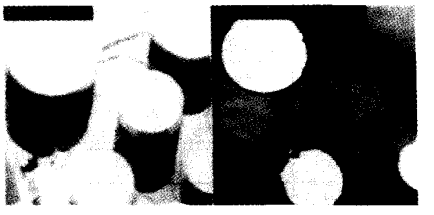
Figure 3 Interpretation of the 3dB SNR Image with an Unknown Object

a. original image b. segmented image

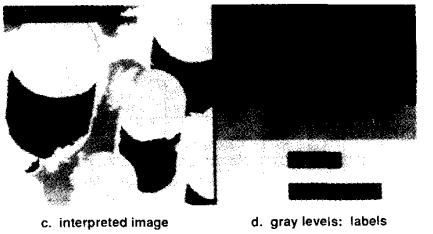


c. interpreted image d. gray levels: labels

Figure 4 Computer Segmentation and Interpretation of Training Image



a. original image b. segmented image



c. interpreted image d. gray levels: labels

Figure 5 Imterpretation of the Test Image

REFERENCES

[1]. T. E. Avery, *Interpretation oJ Aerial Photographs*, Burgess Publishing Company, Minneapolis. Minnesota, 1977.

[2]. S. A. Drury, *Image Interpretation in Geology*, Allen and Unwin, Ltd., London, UK, 1987.

[3] D. P. Paine, *Aerial Photograpl y and Image Interpretation for Resource Management*, John Wil·y and Sons, Inc., New York, 1981.

[4] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1976.

[5] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd Ed., Academic Press, New York, 1982.

[6] B. Horn, *Robot Vision*, MIT P·ess, Cambridge, MA, 1986.

[7] Proc. of the 6th Int. Conf. on Robot Vision and Sensor Control Paris, France, 1986.

[8] R. A. Schowengerdt, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York, 1983.

[9] N. J. Nilsson, *Principles of Artificial Intelligence*, Tioga Pub. Company, Palo Alto, CA, 1980.

[10] R. D. Keller, *Expert System Technology: Development and Application*, Lourdon Press, Englewood Cliffs, NJ, 1987.

[11] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York, 1980.

[12] T. Binford, "Survey of model-based image analysis systems", The Int. J. of Robotics Research, Vol. 1, No. 1, pp. 587-633, 1982.

[13] Y. Ohta, *Knowledge-based Interpretation of Outdoor Natural Color Scenes*, Pitman Advanced Publishing Program, Boston, 1985.

[14] O. D. Faugeras and K. E. Price, "Semantic description of aerial images using stochastic relaxation", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-3, pp. 633-642, Nov., 1981.

[15] K. E. Price, "Relaxation matching techniques: a comparision", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-7, pp. 617-623, Sept., 1985.

[16] V. S. S. Hwang, "Evidence accumulation for spatial reasoning in aerial image understanding", Ph.D. Thesis, Univ. of Maryland, College Park, 1984.

[17] V. S. S. Hwang, T. Matsuyama, L. Davis, and A. Rosenfeld, "Evidence Accumulation for Spatial Reasoning in Aerial Image Understanding", CS-TR-1300, Univ. of Md., 1983.

[18] R. Prasannappa, L. Davis, and V. S. S. Hwang, "A knowledge-based vision system for aerial image understanding", CS-TR-1758, Univ. of Md., Jan., 1987.

[19] R. Brooks, "Symbolic reasoning among 3-dimensional models and 2-dimensional images", *Artificial Intelligence*, Vol. 17, pp. 285-394, 1981.

[20] D. M. McKeown, W. A. Harvey, and J. McDermott, "Rule-based interpretation of aerial imagery", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-7, pp. 570-585, Sept., 1985.

[21] D. M. McKeown and W. A. Harvey, "Automating knowledge acquisition for aerial image interpretation", TR., CMU-CS-87-102, Carnegie-Mellon Univ., Jan., 1987.

[22] E. M. Riseman and A. R. Hanson, "A methodology for the development of general knowledge-based vision systems", TR., COINS 86-27, Univ. of Mass., July, 1986.

[23] A. R. Hanson and E. M. Riseman, "From image measurements to object hypothesis", TR., COINS 87-129, Univ. Mass, Dec., 1987.

[24] B. Draper, R. Collins, J. Brolio, A. R. Hanson and E. M. Riseman, "The schema system", TR., COINS 88-76, Univ. Mass, Sept., 1988.

[25] E. M. Riseman and A. R. Hanson, "Summary of image understanding research at the University of Massachusetts", TR., COINS 88-32, Univ. of Mass., May, 1988.

[26] "Technical reports of the computer vision laboratory, 1986-1988", Univ. of Md., 1988.

[27] J. Zhang, "Merging of segmented regions through simulated annealing", RPI Technical Report, in preparation.

[28] J. Zhang, "Consistent Combination of local interpretation for image analysis", RPI Technical Memo, May, 1988.

[29] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-6, pp. 721-741, Nov., 1984.

[30] C. W. Therrien, T. F. Quatieri and D. E. Dudgeon, "Statistical model-based algorithms for image analysis", Proc. IEEE, Vol. 74, April, 1986.

[31] J. Zhang and J. W. Modestino, "Markov random fields with applications to texture classification and discrimination", Proc. The 20th Annual Conf. on Information Science and Systems, Princeton University, NJ, March, 1986.

[32] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-9, pp. 39-55, Jan., 1987.

[33] F. S. Cohen and D. B. Cooper, "Simple, parallel, hierarchical and relaxation algorithms for segmenting non-casual Markovian random field models", Proc. IEEE Pattern Anal. Machine Intel., Vol. PAMI-9, pp. 195-219, March, 1987.

[34] J. Besag, "On the statistical analysis of dirty pictures", J. Royal Stat. Soc. B., Vol. 48, pp. 259-302, 1986.

[35] J. Zhang and J. W. Modestino, "Unsupervised image segmentation Using a Gaussian model", to be submitted to IEEE Trans. PAMI.

[36] J. Marroquin, S. Mitter, and T. Poggio, "Computer vision", J. Amer. Stat. Association, Vol. 82, pp. 76-89, March, 1987.

[37] P. B. Chou and C. M. Brown, "Multi-modal segmentation using Markov random fields", Proc. IJCAI, pp. 663-670, 1987.

[38] R. Kinderman and J. L. Snell, *Markov Random Fields and Their Applications*, Providence, RI: Amer. Math. Soc., 1980.

[39] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", J. Roy. Statist. Soc , Series B., Vol. 36, pp. 192-226., 1974.

[40] J. A. Feldman and Y. Yakimovsky, "Decision theory and artificial intelligence: I. a semantic-based region analyzer", Artificial Intelligence, Vol. 5, pp. 325-348, 1974.

[41] J. W. Modestino, "A hierarchical region-based approach to automated photointerpretation", RPI report, March 1987.

[42] L. Zadeh, "Approximate Reasoning Based on Fuzzy Logic", Proc. 6th IJCAI, pp.1004-1010, 1979.

[43] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesly Pub. Company, Reading, MA, 1974.

[44] S. Kirkpatrick, C. S. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing", *Science*, Vol. 220, pp. 671-680, May, 1983.

[45] A. A. El Gamal, L. A. Hemachandra, I. Shperling and V. K. Wei, "Using simulated annealing to design good codes", IEEE Trans. Infor. Theory, Vol. IT-33, pp. 116-123, Jan., 1987.

[46] J. Zhang, "Two-dimensional stochastic model-based image analysis", Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, New York, Aug., 1988.

[47] C. S. Won and H. Derin, "Segmentation of noisy textured images using simulated annealing", Proc. ICASSP, pp. 563-566, 1987.

[48] B. Gidas, "Nonstationary Markov chains and convergence of the annealing algorithm", J. Statist. Phys., Vol. 39, pp. 73-131, 1985.

[49] J. Zhang and J. W. Modestino, "Image segmentation using a Gaussian model", Proc. Conf. on Info. Sci. and System, Princeton University, NJ, March, 1988.

[50] J. Zhang, "Learning in the MRF model-based approach for image interpretation", in preparation.

[51] J. Kanai, "Interpretation of real images using the MRF model-based method", RPI report, September, 1987.

[52] L. Y. Yu, "New results on image segmentation and interpretation using the MRF model", RPI report, June, 1988.

[53] C. Bouman and B. Liu, "Segmentation of textured images using a multiple resolution approach", Proc. ICASSP, pp. 1124-1127, New York, March, 1988.

[54] A. Rosenfeld, R.A. Hummel, and S.W. Zucker, "Scene labeling by relaxation operations," IEEE Trans. Syst., Man, Cybern., Vol. SMC-6, pp. 420- 433, 1976.

[55] R.M. Haralick and L.G. Shapiro, "The consistent labeling problem: Part I," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-1, pp. 173-184, Apr. 1979.

[56] R.M. Haralick and L.G. Shapiro, "The consistent labeling problem: Part II," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-2, pp. 193-203, May 1980.

[57] R.A. Hummel and S.W. Zucker, "On the foundation of relaxation labeling process," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-5, pp. 267-287, May 1983.

[58] J.W. Modestino and J. Zhang, "A MRF model-based approach to image interpretation," submitted to IEEE Trans. Pattern Anal. Machine Intel.

[59] J. Zhang, "Implementation of automated photointerpretation approach," 1988 Annual Meeting, Northeast Artificial Intelligence Consortium (NAIC), Technical Report Series, Blue Mountain Lake, NY, August, 1988.

[60] D.E. Rumelhart and J.L. McClelland, eds., *Parallel Distributed Processing*, MIT Press, Cambridge, Mass., 1986.