JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION
Vol. 1, No. 2, November, pp. 158–175, 1990

Toward the Unsupervised Interpretation of Outdoor Imagery

MARK J. CARLOTTO

TASC, 55 Walkers Brook Drive, Reading, Massachusetts 01867

Received February 28, 1990; accepted September 5, 1990

A new approach to the interpretation of outdoor scenes is described. It is based on a context-sensitive classifier which uses relative constraints to describe global relationships between object classes. Contextual models represent the structure of a scene's underlying feature space in terms of stable, physically based parameters. A discrete relaxation algorithm is used to find unambiguous labelings that satisfy a set of ordering relations between object classes. Unlike rule-based systems, these constraints provide a complete and consistent description of the scene. Scenes that are similar in structure are organized into contexts, each of which is represented by a consistent set of constraints. Instead of attempting to achieve a high degree of specificity and localization within limited domains, the methodology is geared toward recognizing general kinds of objects with little or no human intervention over a wider range of scenes. Several examples which demonstrate the recognition of simple objects in black-and-white and multispectral imagery acquired by aircraft, satellite, and at ground level are presented. Through a series of experiments, the ability of the system to degrade gracefully in performance when faced with new and unknown situations is demonstrated.

1. INTRODUCTION

The ability to recognize objects independent of viewpoint and illumination has been a major goal in image understanding research for over 20 years. The greatest success has been achieved in domains that are highly structured and amenable to model-based approaches. (Binford [2] surveys a number of model-based image analysis systems.) Less success has been achieved in interpreting unstructured domains such as general outdoor scenes and aerial imagery. The tremendous diversity in nature makes it difficult if not impossible to account for every possible situation that can occur and to develop precise geometrical models or scene invariant descriptions capable of handling the many sources of variability.

Two ways of dealing with diversity and variability are to partition the domain into distinct kinds of scenes or to develop general purpose models that will work over a wide range of situations. Most of the systems that have been developed to date for understanding outdoor scenes have followed the former approach, in effect dividing the general problem to a series of smaller ones that are more readily handled by model-based approaches. Zucker *et al.* [30] point out that this is not enough; one must develop general purpose models that can deal with unexpected situations. The lack of success to date in this area seems to confirm their point of view.

Fischler and Strat [10] discuss four shortcomings in current machine vision systems with respect to the problem of understanding outdoor scenes: the strong reliance on geometrical information and shape, the ill-defined nature of segmentation, the lack of effective ways to use contextual information, and the inability to effectively bound the search space. They suggest as an intermediate goal the development of a capability for recognizing general kinds of objects in natural settings, in a sense, trading specificity for generality.

A related and important area of research in the remote sensing community is the development of signature extension methodologies. As in machine vision systems, the objective is to develop techniques that will perform well over a wide range of scenes and imaging conditions. Initial approaches were based on an attempt to match clusters in multispectral feature spaces using a multiplicative and additive signature correction model [13, 17]. More recently, signature extension techniques have been developed that use physically-based measurements such as soil brightness, greenness (biomass), and moisture [14, 7]. For example, Hall and Budhwar [11] describe recent developments in using signature-extension techniques for crop classification based on a plant's greenness profile in time.

The objective of the work reported here is to develop a generalizable approach for interpreting outdoor scenes taken from above as in remote sensing and reconnaissance and from ground level as in autonomous land vehicle applications. Instead of attempting to achieve high accuracy in specificity and localization within limited domains, the goal is to be able to recognize general objects with little or no human intervention over a wider range of scenes. An important requirement is that the performance should degrade gracefully outside anticipated operating regimes; i.e., the system must be able to determine whether an unknown situation has been previously

1047-3203/90 $3.00

encountered and be able to provide a reasonable answer if it has not.

The remainder of the paper is organized as follows: Section 2 summarizes a new approach for interpreting outdoor imagery based on the concept of context-sensitive classification. Section 3 describes the major elements of the classifier: a discrete relaxation algorithm that uses constraints between object classes to find unambiguous assignments of classes to objects (labelings), an efficient ordering scheme to reduce the search for unambiguous labelings, a deductive algorithm for deriving constraints from examples, and conflict resolution and context interpolation techniques for handling unknown and unexpected situations. Section 4 presents three case studies to illustrate the potential utility of the context-sensitive classifier in interpreting black-and-white aerial photography, aircraft and satellite multispectral imagery, and black-and-white ground level photography. Areas for future work are discussed in Section 5. A method for deriving texture and spatial information from black-and-white imagery is described in the Appendix.

2. OVERVIEW OF APPROACH

Previous techniques for interpreting outdoor scenes can be divided into model-based approaches which attempt to match structural representations derived from an image to those corresponding to known objects and feature-based approaches which compare measurements derived from an image to those of known objects. The method presented here is a synthesis of the two approaches and involves classifying collections of objects by matching the structure of their underlying feature spaces.

The structure of a distribution of object classes in a feature space Φ is described by a set of partial orderings, or constraints, between the object classes; e.g.,

$$\omega_1(\phi_1) < \omega_2(\phi_1),\ \omega_2(\phi_1) < \omega_3(\phi_1),\ \omega_2(\phi_2) < \omega_3(\phi_2), \text{ and } \omega_1(\phi_3) > \omega_3(\phi_3);$$

or, to pick a simple example,

TREE(brightness)<GROUND(brightness),
GROUND(brightness)<SKY(brightness),
GROUND(vertical-position)<SKY(vertical-position), and
TREE(roughness)>SKY(roughness).

These constraints defined three kinds of object, trees, sky, and ground, in terms of the *relative* values of three features, brightness, roughness (texture), and vertical position.

Context is the relation of an object to its surroundings, i.e., to the other objects in the scene. Usually we think of context in a spatial sense, as relations between objects in image space. In the above example, object classes are defined in terms of how their *features* relate to those of the other object classes. This is a much more general expression of context as the features may be spectral, textural, structural, as well as spatial in nature.

A set of constraints that defines a distribution of object classes in a consistent fashion is known as a *context*. Figure 2-1 shows three feature spaces. Since (a) and (b) describe three objects by the same set of constraints, namely

$$\omega_1(\phi_1) < \omega_2(\phi_1),\ \omega_2(\phi_1) < \omega_3(\phi_1),\ \omega_1(\phi_2) < \omega_3(\phi_2), \text{ and } \omega_3(\phi_2) < \omega_2(\phi_2),$$

they belong to the same context. A different set of ordering relations describes the collection of objects in (c); therefore, it belongs to a different context. More complex topological relationships such as containment, adjacency, and intersection based on representing the extent of an object in feature space are under investigation but are not discussed here.

A collection of objects is classified by a discrete relaxation labeler that attempts to find *unambiguous* labelings which satisfy a given set of constraints by assigning exactly one class to each object. Ideally, if a domain could be divided into contexts that are mutually exclusive and complete, exactly one unambiguous labeling would always be found. When there is overlap between contexts, multiple labelings can occur and *conflict resolution* (selecting the best labeling) is required. When the present situation does not match any known context, *context interpolation* (synthesizing a plausible labeling) may be necessary. Conflict resolution and context interpolation strategies are discussed in Section 3.

The proposed approach differs from previous outdoor scene understanding efforts in three fundamental ways: in the use of context, in the representation of object classes, and in the generation of constraints. Some scene interpretation systems that use context can be described

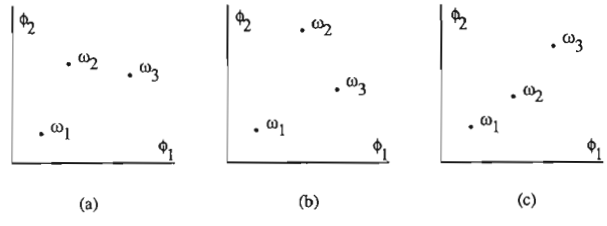


FIG. 2-1. Three feature spaces. A representation of feature spaces using relative constraints is invariant with respect to distortions of the form $f(\phi_m)$, where f is an order-preserving function.

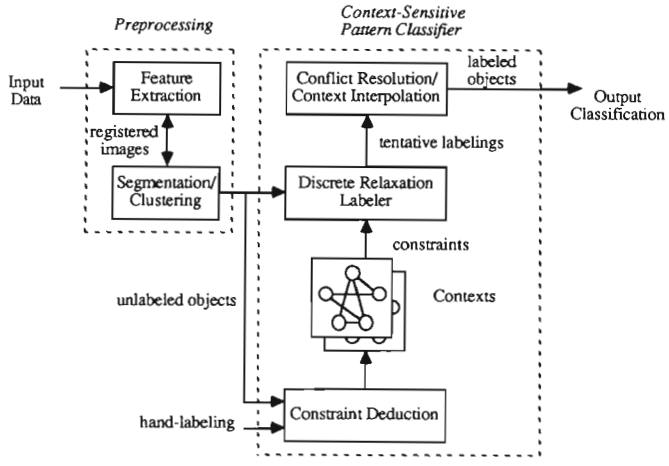


FIG. 2-2. General context-sensitive pattern recognition system.

as context-dependent, i.e., the context must be specified a priori so that the proper schema or knowledge sources can be applied to the data [22]. Here, context is seen as a collective and emergent property of feature spaces so the system is sensitive to, but not dependent on, context in the sense that it must be specified explicitly. If all the contexts are applied to the data being classified only those that are relevant will match or contribute consistent information. Collections of objects are represented by a set of mutually consistent constraints which define partial orderings between objects in terms of a set of underlying physically based features. This differs from the use of context sets [10] where context is used only to enforce consistency between classes during the interpretation process. A number of systems have been developed for interpreting outdoor scenes that use constraint-based representations [1, 26] or production systems [19, 21]. Here, classification involves finding unambiguous labelings. Moreover, the classifier uses constraints that are consistent and complete which can be derived automatically from examples through a simple deductive procedure as described in Section 3. It is noted that the proposed method is similar in certain ways to the representation and matching of pictorial structures developed by Fischler and Elschlager [9]. Here the compatibility between hypotheses is treated as a cost function that is used by the discrete relaxation labeler for prioritizing the search for unambiguous labelings [4]. Fischler and Elschlager use a dynamic programming algorithm that finds low cost solutions. It is shown that unambiguous labelings do not always have the lowest cost.

Section 3 describes in detail all the major elements of the context-sensitive classifier. As shown in Fig. 2-2, the input to the classifier is a set of unlabeled objects that represent significant modes or clusters in the feature space. A statistical clustering algorithm is used here to segment one or more registered images by decomposing the corresponding feature space into a mixture of multivariate normal components. Depending on the domain and the sensor, different feature extraction techniques are required to create a registered set of input images. In order to recognize surface materials in multispectral imagery (Section 4.1), the tasselled cap brightness and greenness are computed from the reflective bands of the sensor. In other domains (Sections 4.2 and 4.3) texture and spatial information must be derived from a single image. One such method that has been developed for computing registered representations of image brightness, texture, and spatial information is described in the Appendix.

3. CONTEXT-SENSITIVE PATTERN CLASSIFICATION

This section describes major elements of the context-sensitive classifier: a discrete relaxation labeler, an efficient ordering scheme for reducing search, an algorithm for deriving constraints from examples, and methods for conflict resolution and context interpolation.

3.1. *Discrete Relaxation Labeling*

The use of discrete relaxation labeling techniques can be traced back to Waltz. The problem addressed was that of interpreting line drawings of 3-D objects using knowledge about the compatibility of line junctions. Feldman and Yakimovsky [8] developed a semantics-based region analyzer which segmented and labeled images of simple scenes. Tenenbaum and Barrow [28] developed an interpretation-guided segmentor similar in concept to the one above based on Waltz's filtering ideas. Rosenfeld *et al.* [25] formalized the theory of discrete, fuzzy, and probabilistic relaxation methods.

Let $A = \{a_1, a_2, \ldots, a_N\}$ be the set of objects we wish to classify, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$ be the set of possible labels or classes for the objects, and $\Phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$ be the set of properties defined over the set of objects. The value of the mth property of the nth object is denoted $a_n(\phi_m)$. Although we do not refer directly to the numerical values of classes, the notation $\omega_k(\phi_m) > \omega_{k'}(\phi_m)$ means that in terms of the mth property, the kth class is strictly greater than the k'th class. The set of hypotheses, $H = A \times \Omega = \{h_{nk}\}$, represents all possible pairings of objects and labels; h_{nk} is the hypothesis that object a_n is a member of class ω_k. $R = \Phi \times \Omega \times \Omega$ is a set of constraints where $r_{mkk'} \in \{0, 1\}$. If $\omega_k(\phi_m) > \omega_{k'}(\phi_m)$ then $r_{mkk'} = 1$ (true); or if $\omega_k(\phi_m) \leq \omega_{k'}(\phi_m)$ then $r_{mkk'} = 0$ (false); or if, in terms of ϕ_m, there is no ordering relation between ω_k and $\omega_{k'}$, $r_{mkk'} = \varnothing$ (not defined). By definition, $r_{mkk} = \varnothing$. As a result there may be at most $MK(K-1)/2$ constraints, although some may be redundant, e.g., $r_{mkk'} = 1$ and $r_{mk'k''} = 1 \rightarrow r_{mkk''} = 1$. If the two

classes ω_k and $\omega_{k'}$ cannot be distinguished from one another in terms of the $\{\phi_m\}$, $r_{mkk'} = \varnothing$ for all m.

Let $\delta[h_{nk}, h_{n'k'}]$ denote the *compatibility* of hypotheses h_{nk} and $h_{n'k'}$. Two hypotheses are compatible, i.e., $\delta = 1$ if

(i) $r_{mkk'} = \varnothing$ for all m, or

(ii) $n \neq n'$ and $k = k'$, since two objects may belong to the same class, or

(iii) $n = n'$ and $k \neq k$, since an object may belong to more than one class, or

(iv) $n = n'$ and $k = k'$, since a hypothesis is compatible with itself, or

(v) for each $r_{mkk'}$, $a_n(\phi_m) > a_{n'}(\phi_m)$ since a_n is associated with ω_k, $a_{n'}$ is associated with $\omega_{k'}$, and $r_{mkk'}$ requires that $\omega_k(\phi_m) > \omega_{k'}(\phi_m)$;

otherwise $\delta = 0$. (i) is the case where there are no constraints between the classes; (ii) is possible in situations where the number of objects is greater than the number of classes; (iii) is possible during the initial phases of the labeling process (i.e., during Waltz filtering); (iv) is the trivial case. In (v), if any constraint is violated, then h_{nk} and $h_{n'k'}$ are not compatible.

A *labeling* is an assignment of classes to objects. A *consistent* labeling consists of all hypotheses h_{nk} that are compatible with at least one $h_{n'k'}$ $n \neq n'$ and $k \neq k'$. In a consistent labeling each object may have more than one label. The process of finding consistent labelings, termed Waltz filtering, involves repeatedly applying constraints to hypotheses, eliminating hypotheses which are not compatible with at least one other hypothesis, until the process converges. Rosenfeld *et al.* [25] proved that this process always converges. An *unambiguous* labeling assigns exactly one class or label per object. If there is more than one way to do this, multiple unambiguous labelings may be obtained. If there is no way to do this one can select, for each object, the class corresponding to the most compatible hypothesis as is discussed in Section 3.3.

Ultimately, we are interested in finding unambiguous labelings. Exhaustive search for unambiguous labelings, e.g., via the tree search procedure developed by Waltz,

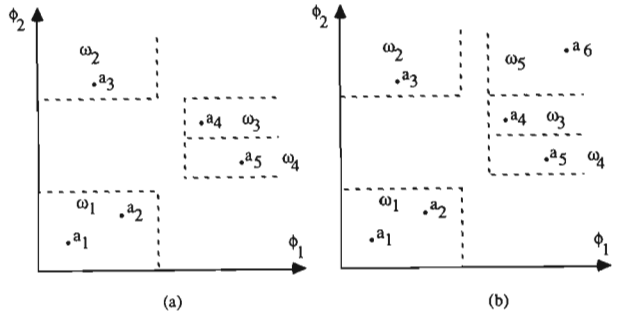


FIG. 3-1. Two distributions of objects in a 2-D feature space.

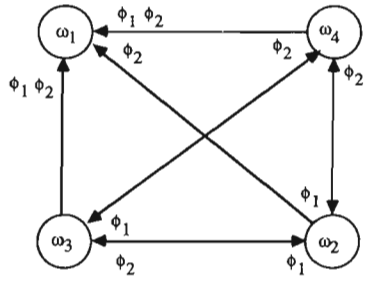


FIG. 3-2. Directed graph representation of constraints.

can require the examination of up to K^N labelings. Unary constraints (which involve one object at a time) can, in conjunction with Waltz filtering, be used to prune large portions of the search tree by eliminating incompatible hypotheses at the top of the tree. Here, constraints are binary (involve two objects at a time) and classes are only weakly constrained relative to one another. Typically, there are few incompatible hypotheses and so Waltz filtering is not effective in itself in reducing search.

3.2. *Classification Using Relative Constraints*

Figure 3-1a shows a distribution of clusters (a_1 through a_5) in a two-dimensional feature space. The clusters represent a possible segmentation of an image and are defined by two properties: ϕ_1 and ϕ_2. Initially, we assume that four classes $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ are present in the image. Figure 3-2 is a directed graph representation of these four classes in terms of only relative constraints. For example, the top arc represents the two constraints $\omega_4(\phi_1) > \omega_1(\phi_1)$ and $\omega_4(\phi_2) > \omega_1(\omega_2)$ ($r_{141} = 1$ and $r_{241} = 1$). Figure 3-1a also shows a labeling that satisfies the constraints in Fig. 3-2. Since it assigns one class per cluster, it is an unambiguous labeling. Qualitatively, unambiguous labelings will be found only if the structure of the feature space matches the structure defined by the constraints. The dotted lines in Fig. 3-1a are the decision boundaries (parallelpipeds in higher dimensional spaces) induced by the constraints in Fig. 3-2. Their relationships to one another are defined by the constraints, but their numerical values depend on the data being classified. Relative constraints capture the order of the data in each dimension and are therefore insensitive to distortions of the form $a_m\phi_m + b_m$, where a_m and b_m are unknown constants and $a_m > 0$. It is noted that certain signature extension techniques [17, 13] are based on a model of the image formation process that attempts to account for and correct differences between images by additive and multiplicative factors. Here, such corrections are not necessary.

Figure 3-3 is a search tree for the above example and contains $4^5 = 1024$ paths, each of which corresponds to a labeling. The dotted lines in the figure correspond to the labelings that are eliminated by Waltz filtering. For this

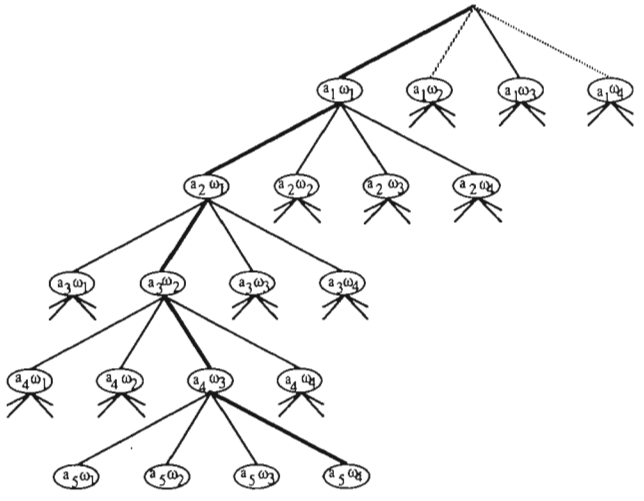


FIG. 3-3. Hypothesis tree for depth-first search for unambiguous labelings.

example, Waltz filtering eliminates 2 out of 20 possible hypotheses, which given the ordering of the hypotheses, reduces the number of candidate labelings by 50%.

An even more effective way to reduce search is to order hypotheses by their *mutual incompatibility* and to visit those hypotheses first. The mutual incompatibility of hypothesis h_{nk} is

$$c_{nk} = \sum_{n' \neq n, k' \neq k} \delta[h_{nk}, h_{n'k'}].$$

Table 3-1 lists, for each of the 18 hypotheses h_{nk} which survived Waltz filtering, a *cost* that is equal to their mutual incompatibility. If a hypothesis is compatible with every other hypothesis in H, its cost is zero. In each row, the most compatible hypothesis (lowest cost) is shown in boldface. The heuristic is that unambiguous labelings will contain low cost hypotheses, and conversely, labelings consisting of hypotheses with low costs are likely to be unambiguous. The hypotheses which belong to the unambiguous labeling in Fig. 3-1a are marked with asterisks in Table 3-1. In this particular example, if the tree is built such that the most compatible hypotheses are examined first, the correct labeling is found immediately as it has the lowest cost.

TABLE 3-1
Hypothesis Table for Objects in Figure 1a

Object	Class ω_1	ω_2	ω_3	ω_4
a_1	**0***	8		
a_2	**3***	7	8	7
a_3	7	**0***	6	10
a_4	8	5	**5***	6
a_5	8	8	5	**3***

Note. The most compatible assignments are in boldface.*

TABLE 3-2
Hypothesis Table for Objects in Figure 3a

Object	Class ω_1	ω_2	ω_3	ω_4
a_1	**0**	9		
a_2	**3**	8	10	9
a_3	7	**1**	8	12
a_4	8	**6**	7	8
a_5	8	9	7	**5**
a_6		8	**4**	9

Note. No unambiguous labelings exist.

Consider now a second example. Figure 3-1b shows the five clusters from the previous example plus a sixth. If we try to label all six clusters in Fig. 3-1b using the constraints in Fig. 3-2, no unambiguous labeling exists (Table 3-2). If a fifth class and four additional constraints are added, namely, $\omega_5(\phi_1) > \omega_4(\phi_1)$, $\omega_3(\phi_1)$, $\omega_2(\phi_1)$, and $\omega_1(\phi_1)$, the correct unambiguous labeling is obtained (Table 3-3). It is worth noting however that in this example, the hypotheses which belong to the unambiguous labeling are not always the ones with the lowest cost. Thus, although searching through the most compatible hypotheses provides a good starting point, some backtracking may be needed.

3.3. *Deducing Constraints from Examples*

A mutually consistent and complete set of constraints can be derived from examples using a simple deductive procedure motivated by Valiant [29]. Initially, assume that all $r_{mkk'} = \{0, 1\}$ which implies $\omega_k(\phi_m) > \omega_{k'}(\phi_m)$ and

TABLE 3-3
Hypothesis Table for Objects in Figure 3a with Added Constraints for ω_5

Object	Class ω_1	ω_2	ω_3	ω_4	ω_5
a_1	**0***	9	14	14	
a_2	**6***	11	12	11	11
a_3	9	**3***	9	26	15
a_4	12	10	10*	11	**7**
a_5	13	14	11	9*	**3**
a_6		14	9	14	**0***

Note. The hypotheses in the unambiguous labeling are not always the most compatible ones.

$\omega_k(\phi_m) \leq \omega_{k'}(\phi_m)$ for all k, k', and m. Since $r_{mkk'} = \neg r_{mk'k}$ only $MN(N-1)/2$ constraints are involved. Next, for all pairs of clusters, a_n and $a_{n'}$ that belong to classes ω_k and $\omega_{k'}$, respectively, if any $a_n(\phi_m) > a_{n'}(\phi_m)$, we eliminate the "0" from the corresponding $r_{mkk'}$. If any $a_n(\phi_m) \leq a_{n'}(\phi_m)$, we eliminate the "1" from the corresponding $r_{mkk'}$. This is done for all pairs of classes and for all properties. At the end, if $r_{mkk'} = \{0, 1\}$ or is null, there is no constraint between ω_k and $\omega_{k'}$ along ϕ_m. Table 3-4 shows the earlier example of deriving constraints for the labeled objects in Fig. 3a. All relevant entries in the table are initially set to {0, 1}. In (b), the effect of adding the interaction of hypotheses h_{21} and h_{32} to the table is seen. (Hypotheses h_{21} and h_{32} correspond to assigning class ω_1 to object a_2 and class ω_2 to object a_3.) The final state of the table is shown in (c).

TABLE 3-4
Deductive Learning of Constraints by Manually Assigning Labels to Objects

(a) Initial constraint table, $r_{mkk'}$

		ω_1		ω_2		ω_3		ω_4	
		ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2
ω_1	ϕ_1								
	ϕ_2								
ω_2	ϕ_1	{0,1}							
	ϕ_2		{0,1}						
ω_3	ϕ_1	{0,1}		{0,1}					
	ϕ_2		{0,1}		{0,1}				
ω_4	ϕ_1	{0,1}		{0,1}		{0,1}			
	ϕ_2		{0,1}		{0,1}		{0,1}		

(b) Updated with the $\{h_{21}, h_{32}\}$ hypothesis pair

		ω_1		ω_2		ω_3		ω_4	
		ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2
ω_1	ϕ_1								
	ϕ_2								
ω_2	ϕ_1	∅							
	ϕ_2		{1}						
ω_3	ϕ_1	{0,1}		{0,1}					
	ϕ_2		{0,1}		{0,1}				
ω_4	ϕ_1	{0,1}		{0,1}		{0,1}			
	ϕ_2		{0,1}		{0,1}		{0,1}		

(c) Final state after all hypothesis pairs have been accumulated

		ω_1		ω_2		ω_3		ω_4	
		ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2
ω_1	ϕ_1								
	ϕ_2								
ω_2	ϕ_1	∅							
	ϕ_2		{1}						
ω_3	ϕ_1	{1}		{1}					
	ϕ_2		{1}		{0}				
ω_4	ϕ_1	{1}		{1}		{1}			
	ϕ_2		{1}		{0}		{0}		

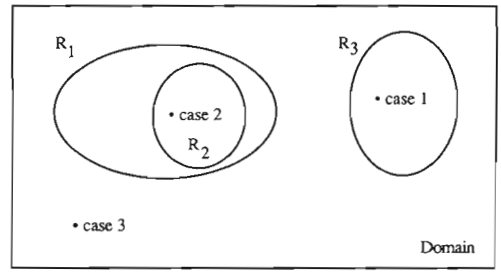


FIG. 3-4. Domain partially covered by three contexts R_1, R_2, and R_3. Three cases corresponding to one labeling (1), more than one labeling (2), and no labeling (3) are shown.

3.4. *Conflict Resolution and Context Interpolation*

Ideally, if a domain could be divided into contexts that are mutually exclusive and complete at least one unambiguous labeling would always be found for each situation encountered. (As discussed below, when there are more objects than classes and the classes are not sufficiently constrained, multiple unambiguous labelings can occur.) In reality, more specific contexts may be contained within less specific ones and the domain may not be completely covered as depicted in Fig. 3-4. When there is overlap between contexts, i.e., when multiple constraint sets are applied to the data and two or more of the constraint sets produce unambiguous labelings, *conflict resolution* (selecting the best labeling) is required. When the present situation does not match any known context, some form of *context interpolation* (synthesizing a plausible labeling) may be necessary.

The incompatibility of a labeling $C(\Lambda)$ is defined as the sum of the mutual incompatibilities of the hypotheses which comprise the labeling, i.e.,

$$C(\Lambda) = \sum_{n,k} c_{nk}\lambda_{nk},$$

where $\lambda_{nk} = 1$ if class ω_k is assigned to object a_k and zero otherwise. Thus $C(\Lambda)$ is the total cost of the labeling. When one set of constraints yields multiple unambiguous labelings, the one with the lowest cost is selected. It is shown later in Section 4.1 that the correct labeling is the unambiguous labeling with the lowest cost.

Conflicts between labelings from different contexts are resolved by picking the labeling that corresponds to the most specific situation. More general contexts can be created by merging similar classes (e.g., trees and grass into vegetation). This is desirable from the point of view of being able to provide some kind of answer when an unexpected situation is encountered. However, if more and less specific contexts are used, together a means for selecting the proper labeling in the event of a conflict is necessary. Picking the most specific labeling is based on

the observation that more specific contexts can usually be reduced to less specific ones by combining classes and eliminating constraints. For example, the context defined by the constraints

$$\omega_2(\phi) > \omega_1(\phi),\ \omega_1(\phi) > \omega_3(\phi),\ \omega_4(\phi) > \omega_3(\phi), \text{ and } \omega_4(\phi) > \omega_2(\phi)$$

can be reduced to a more general context by combining ω_1 and ω_3 into a new class ω_5. This generalized context is defined by the relations

$$\omega_2(\phi) > \omega_5(\phi),\ \omega_4(\phi) > \omega_5(\phi), \text{ and } \omega_4(\phi) > \omega_2(\phi).$$

It can be shown that both contexts will produce unambiguous labelings. An example that illustrates the conflict resolution strategy is provided in Section 4.2.

While generalization provides a way of increasing coverage of the domain, there will be new situations where no unambiguous labeling will be found and one would like to "interpolate" information from relevant contexts. A problem is how to decide what contexts are relevant. One approach is to generate a set of tentative labelings (one for each context) by picking the most compatible hypothesis for each object. These labelings are not globally consistent and so certain objects will be misclassified in each one. It is argued that if there are enough contexts, the contexts that are relevant will tend to vote for the same or, at least similar, classes while those that are irrelevant will vote for different classes. Therefore a reasonable strategy is to choose the class with the most votes. An example is presented in Section 4.2 that illustrates this idea.

4. EXPERIMENTAL RESULTS

This section presents three case studies that involve identifying general kinds of objects in imagery. Several domains are considered: black-and-white aerial photography, aircraft and satellite multispectral imagery, and ground level black-and-white photography.

4.1. *Surface Feature Classification Using Texture*

The focus in the first case study is on the classification of general surface features such as trees, shadows, sparse vegetation, and bare soil in black-and-white imagery. Initially we are interested in assessing the performance of the context-sensitive classifier on its training set. A single image is segmented and clustered into representative areas based on image brightness and texture. The areas are hand labeled and used to generate a set of constraints according to the procedure described in Section 3.3. These constraints are then used to classify the image.

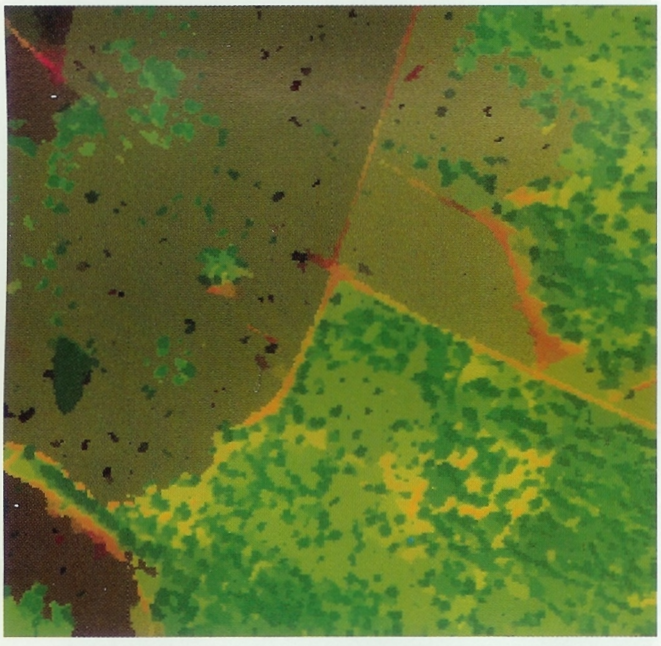
Figure 4-1a depicts brightness (BR) and fractal dimension (FD) images derived from a black-and-white aerial photograph in red and green, respectively. The FD image represents a measure of the local roughness that has been averaged within areas of similar gray-level in the BR image. The algorithms are described in greater detail in the Appendix. A nonsupervised clustering algorithm [12] is then used to segment the image pair into disjoint areas with similar statistics. The algorithm is based on a multivariate Gaussian mixture model and is generalization of an earlier technique that used a scale-space approach for decomposing a histogram into a collection of normal modes [3]. Table 4-1 lists the 10 clusters extracted by the clustering algorithm. A set of constraints was derived by hand labeling these clusters. The directed graph (digraph) representation of these constraints is shown in Fig. 4-2. Trees and bare-soil are rougher than sparse-vegetation and water, water is darkest, sparse-vegetation is least rough, and bare-soil is brightest.

By using these constraints to classify the original clusters using the discrete relaxation labeling algorithm, two unambiguous labelings are obtained (Table 4-1). If two or more unambiguous labelings are obtained from a given set of constraints, the most compatible labeling is picked according to the conflict resolution strategy outlined in Section 3. The first labeling is the most compatible and corresponds to the true (hand-labeled) classification. Figure 4-1b shows the classification image which corresponds to the first labeling. The second labeling was less compatible, misclassifying a cluster corresponding to several areas of bare soil with some brush as trees.

4.2. *Multispectral Image Classification and Signature Extension*

In this case study, several images over different areas are examined. The intent is to study the potential utility of the method for multispectral signature extension, e.g., for building thematic maps over large areas without human supervision. In addition we wish to understand how the performance of the classifier degrades when unexpected situations occur. Both conflict resolution and context interpolation strategies are demonstrated by way of example.

Two images were initially considered: W1 over suburban Washington, DC, and F2 over rural Arkansas. Both were acquired by NASA airborne thematic mapper (ATM) sensors. As shown in Fig. 4-3 (top) the images appear to be quite different in the visible bands. The tasselled-cap brightness (BR) and greenness (GR) images are shown in Fig. 4-3 (middle) in red and green, respectively. Man-made features, dry soil, and concrete have relatively high brightness values and appear red; vegetation is green. For thematic mapper sensors, the bright-



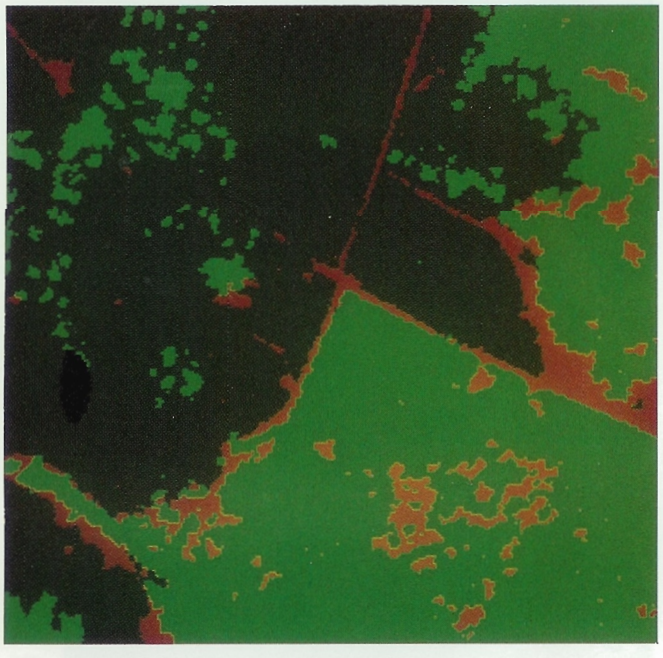


FIG. 4-1. Brightness and fractal dimension (texture) images shown in red and green (top). Classification map (bottom) shows water (black), trees (bright green), sparse-vegetation (dark green), and bare-soil (red).

ness and greenness are given by [7]

$$\begin{aligned} \mathrm{BR} &= .3037x_1 + .2793x_2 + .4743x_3 + .5585x_4 \\ &\quad + .5082x_5 + .1863x_7 \\ \mathrm{GR} &= -.2848x_1 - .2435x_2 - .5436x_3 + .7243x_4 \\ &\quad + .0840x_5 - .1800x_7, \end{aligned}$$

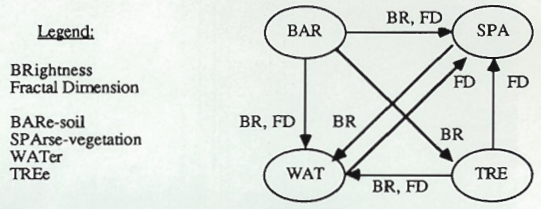


FIG. 4-2. Digraph representation of classes in terms of brightness and fractal dimension.

where the x_i are the brightness values of the six reflective TM bands. Labeled feature spaces for the two scenes are shown in Fig. 4-4. Radiometrically, there are large differences between the two scenes. Structurally, the two feature spaces are quite similar with the exception that sparse vegetation is greener than trees in W1 while the opposite is true in F2.

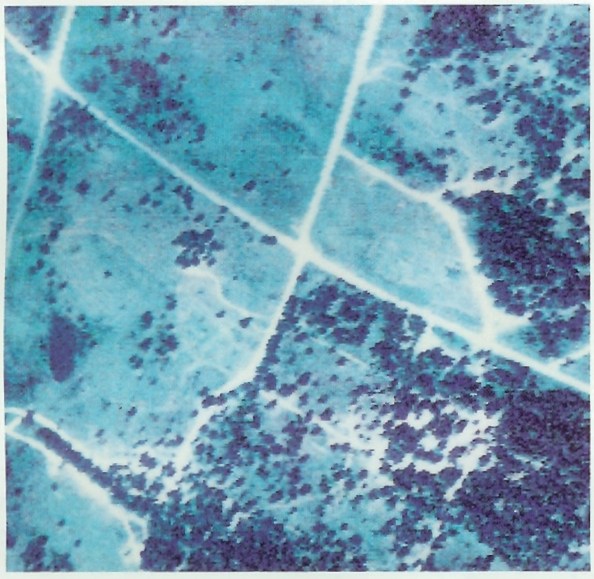
Correct unambiguous labelings are obtained by classifying the training set as before; here, F2 constraints were used to classify F2 and W1 constraints to classify W1 (Fig. 4.3, bottom). We then attempted to classify W1 using F2 constraints, and vice versa. No unambiguous labelings are found since the feature spaces are slightly different in structure. However, if sparse vegetation and trees are lumped together into a more general vegetation category (VEG) and the constraints modified accordingly, the results in Tables 4-2 and 4-3 are obtained.

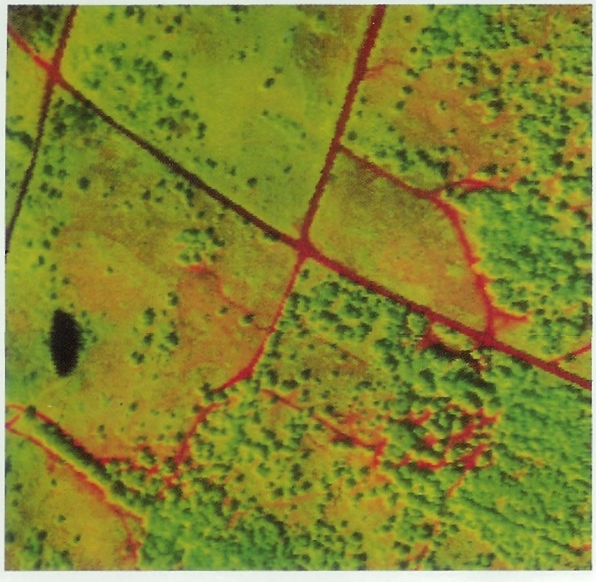
This example shows that the expense of reduced specificity, both scenes can be classified by the same set of constraints. By adding this third context (W1F2) and repeating the experiment using all three context sets, two labelings will be found for each scene. Picking the more specific labeling (i.e., W1 for W1 and F2 for F2) will yield the correct answer. But now if a new situation is encountered that does not correspond to either W1 or F2, but does contain vegetation, bare-soil, and shadows, W1F2 will be able to provide a general answer in lieu of more detailed knowledge.

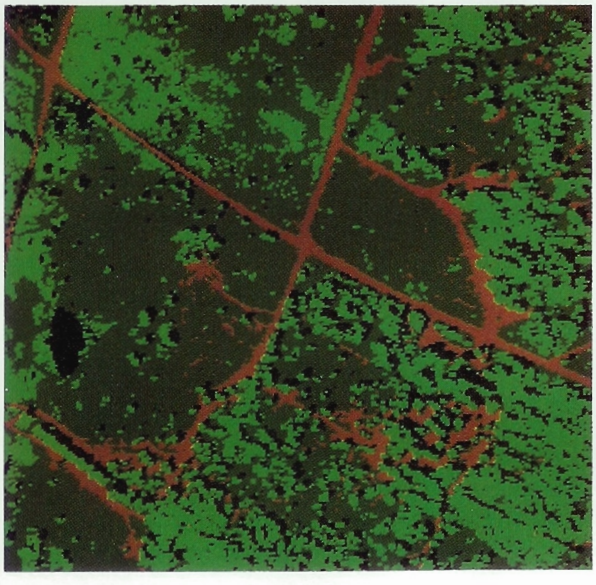
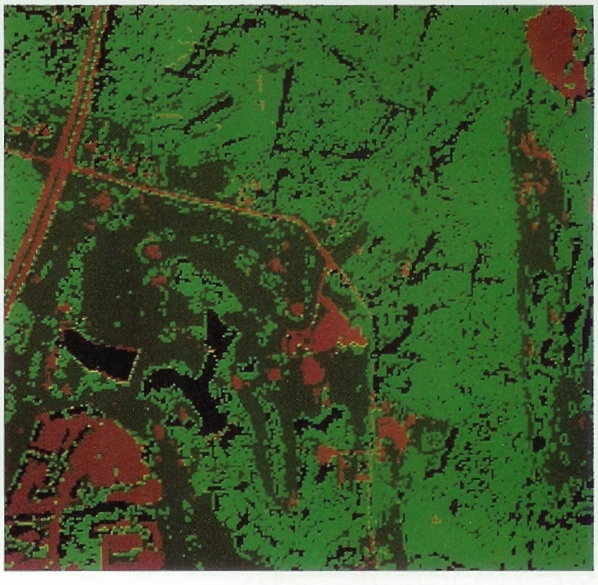
The performance of the context-sensitive classifier can be compared to a statistical technique that matches clusters using a similarity measure based on discrimination information [16]. The discrimination information between two clusters a_n and $a_{n'}$ is given by

$$\begin{aligned} I(n, n') = \sum_m \{&[1/\sigma^2(n, m) - 1/\sigma^2(n', m)][\sigma^2(n, m) \\ &- \sigma^2(n', m)] + [1/\sigma^2(n, m) \\ &+ 1/\sigma^2(n', m)][\mu^2(n, m) - \mu^2(n', m)]\}, \end{aligned}$$

where $\mu(n, m)$ and $\sigma(n, m)$ are the mean and standard deviation for the mth feature of the nth cluster. In Table 4-2 the cluster matching results are obtained by finding,





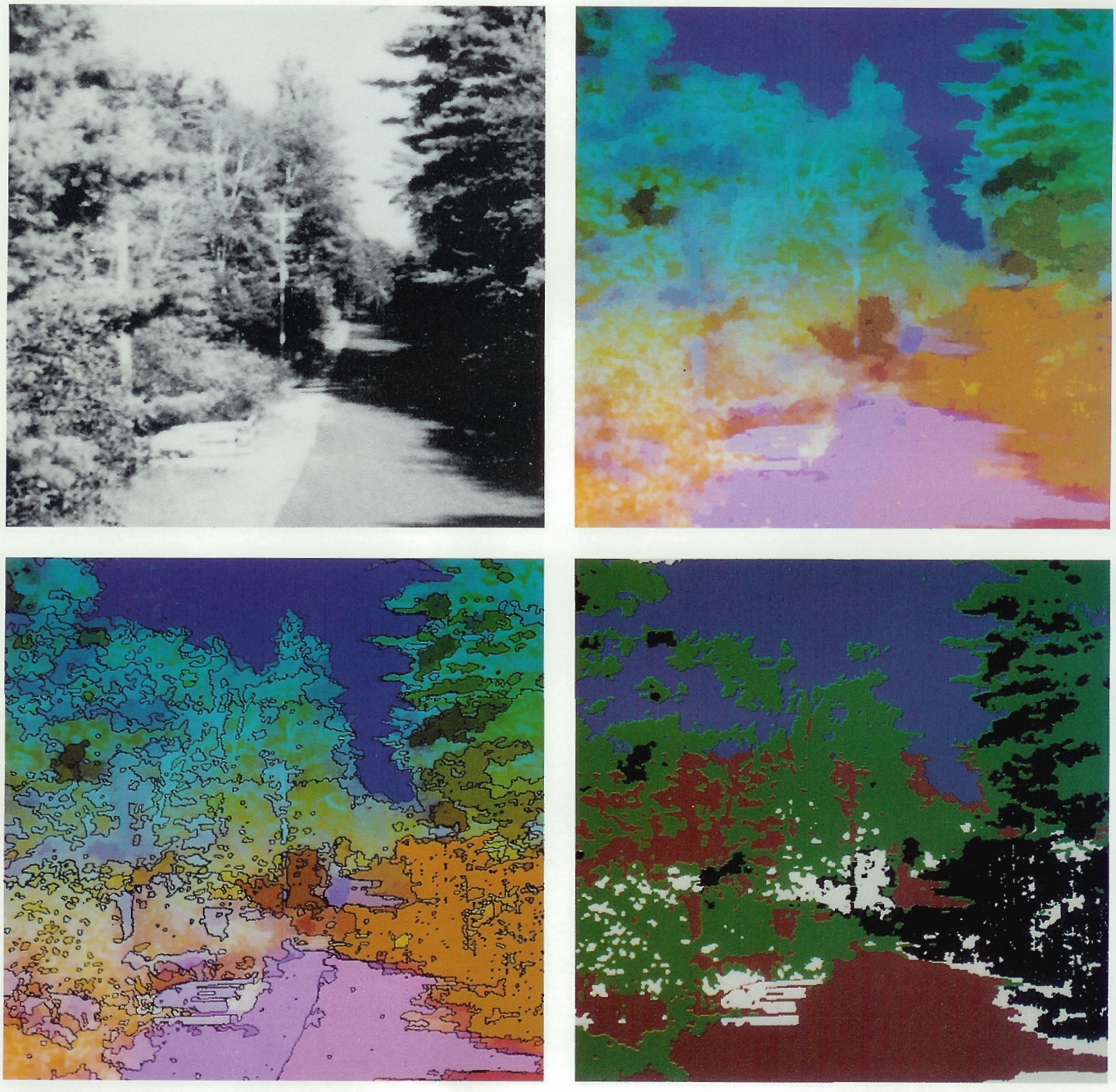


FIG. 4-9. Analysis of S1 image. Original black-and-white image (upper left). False-color rendition depicting brightness in red, fractal dimension in green, and vertical position in blue (upper right). Segmentation overlayed on the false-color image (lower left). Classification image showing trees in green, shadows in black, sky in blue, ground in red, and areas not classified in white (lower right).

FIG. 4-3. At the top, almost true color renditions of images F2 (left) and W1 (right). Spectral bands 1, 2, and 3, (0.45–0.52, 0.52–0.60, and 0.63–0.69 μm) are shown in blue, green, and red. In the middle, brightness and greenness images computed for F2 (left) and W1 (right) using the tasselled-cap transform. At the bottom, classification images for F2 (left) and W1 (right). Trees are green, sparse-vegetation is dark green, bare-soil is red, and water and shadows are black.

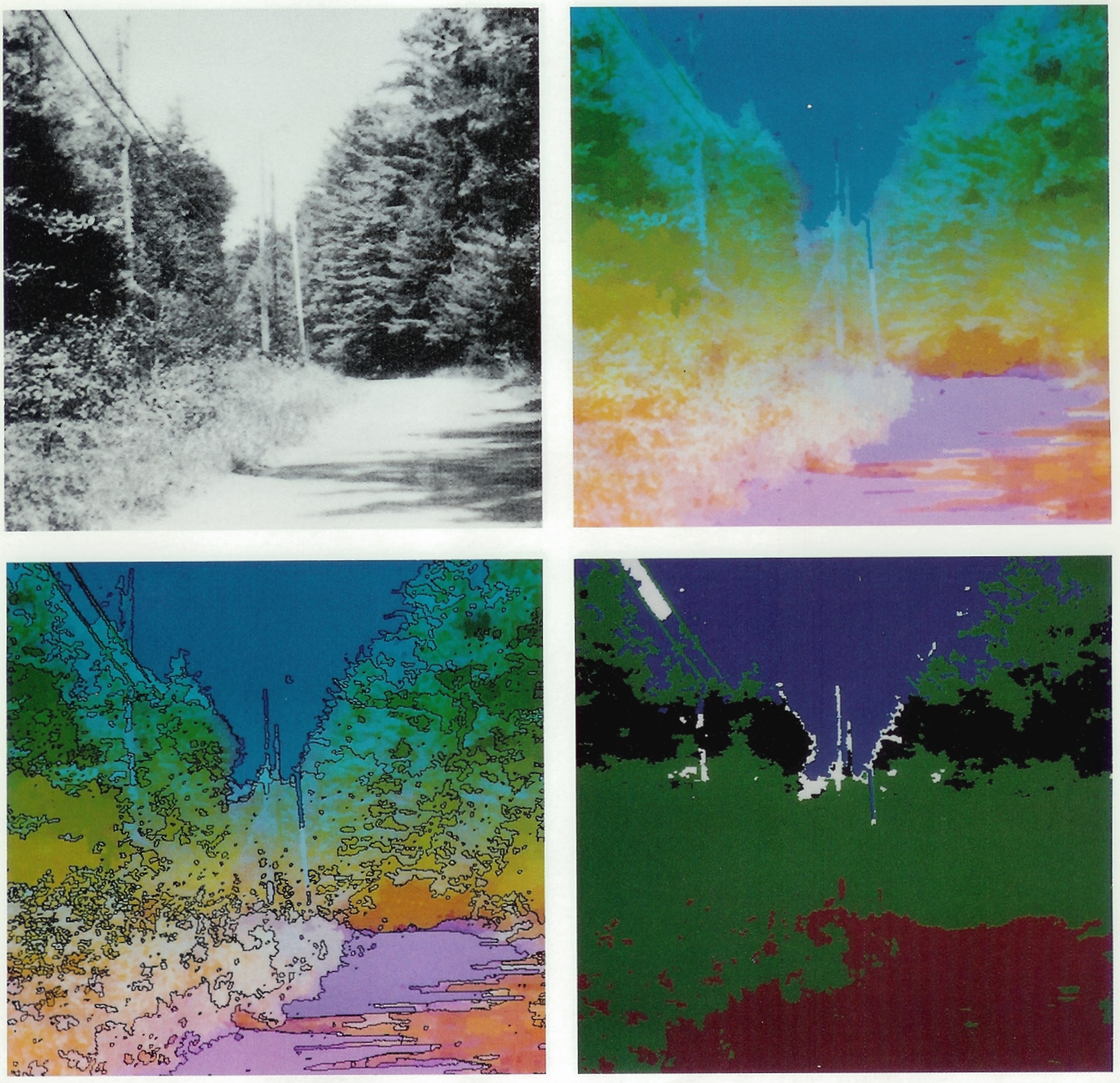


FIG. 4-10. Analysis of S2 image. Original black-and-white image (upper left). False-color rendition depicting brightness in red, fractal dimension in green, and vertical position in blue (upper right). Segmentation overlayed on the false-color image (lower left). Classification image showing trees in green, shadows in black, sky in blue, ground in red, and areas not classified in white (lower right).

for each cluster in W1, the cluster in F2 that has the smallest discrimination information and assigning the class associated with that cluster in F2 to the corresponding cluster in W1. In Table 4-3, the situation is reversed. These results suggest that information about relationships between objects is important in attempting to associate objects (clusters) in one scene to those in another.

In order to examine the extendibility of the representation, three context sets, one for W1, and two others derived from agricultural scenes in the central U.S. and in

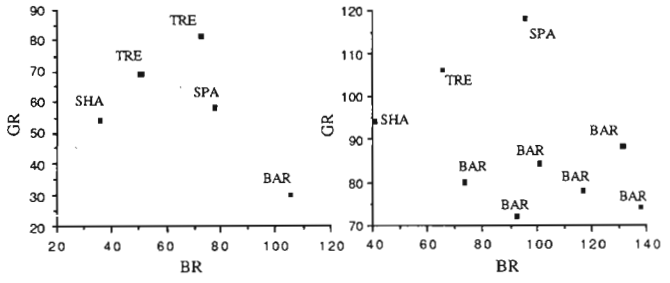


FIG. 4-4. Labeled feature spaces for F2 (left) and W1 (right).

Europe were used to classify F2. The context set for F2 was withheld. KN2 and M2 are Landsat TM images. In effect we are simulating the performance of the context-sensitive classifier to an unknown situation. Figures 4-5 through 4-8 show the digraph representation of F2, along with W1, KN2, and M2. The results of classifying F2 using constraints from KN2, M2, and W1 are summarized in Table 4-4. No unambiguous labelings are found which indicates that F2 does not fit any of the three contexts.

A question is whether one can "interpolate" information from the other contexts. In each labeling shown in Table 4-4, the most compatible hypothesis was selected for each cluster. Each labeling thus contains certain errors. The classification in the last column was derived by majority rule. Although the resultant labeling still contains a few errors, it represents a reasonable response to F2 given the available knowledge. All four scenes contained shadows and water, bare-soil, sparse-vegetation, trees. Shadows and water were combined due to the difficulty in separating water from shadows in most of the imagery. F2 was over an area in Arkansas with little background vegetation so sparse-vegetation was similar to bare-soil in the other scenes. One of the tree clusters in F2 was misclassified as crops since in two of the scenes crops had the greatest biomass.

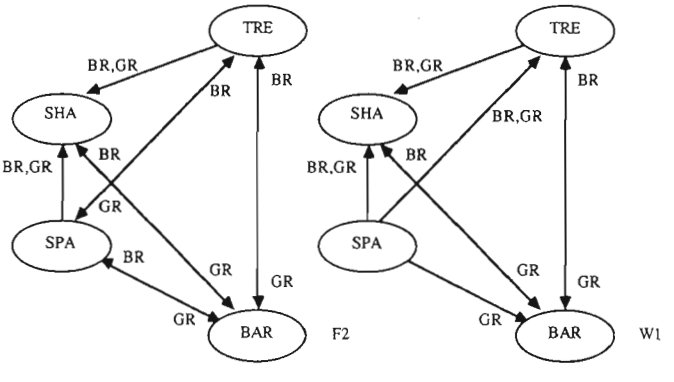


FIG. 4-5. Digraph representation of F2 context.

FIG. 4-6. Digraph representation of W1 context.

TABLE 4-1
Texture Classification Results for Aerial Black-and-White Images

Cluster number	Features		Classification		
	BR	FD	True class	Labeling 1	Labeling 2
0	0	121	WAT	WAT	WAT
1	65	102	SPA	SPA	SPA
2	61	152	TRE	TRE	TRE
3	101	83	SPA	SPA	SPA
4	104	115	SPA	SPA	SPA
5	134	128	BAR	BAR	**TRE**
6	151	208	BAR	BAR	BAR
7	40	200	TRE	TRE	TRE
8	95	107	SPA	SPA	SPA
9	89	198	TRE	TRE	TRE

Note. Misclassification is shown in boldface.

4.3. *Black-and-White Scene Analysis*

The last case study examines the utility of the context-sensitive pattern classifier for interpreting typical outdoor scenes acquired at ground level. Figures 4-9 and 4-10 show two such scenes (S1 and S2). Both are digitized photographs taken in upstate New York. In each, the original black-and-white image is shown in the upper left. A false-color rendition depicting brightness in red, fractal dimension in green, and vertical position (represented region-by-region as is done for brightness and fractal dimension) in blue is shown in the upper right. The images were computed by the method described in the Appen-

TABLE 4-2
Suburban Washington Classification Results Using Brightness and Greenness Features

Cluster	True classification	Unambiguous labeling	Best match using discrimination information
0	SHA	SHA	**TRE**
1	BAR	BAR	**TRE**
2	TRE	VEG	TRE
3	BAR	BAR	**TRE**
4	BAR	BAR	**TRE**
5	SPA	VEG	**TRE**
6	BAR	BAR	BAR
7	BAR	BAR	**TRE**
8	BAR	BAR	**TRE**

Note. Unambiguous labeling obtained by collapsing rules for trees and sparse-vegetation into a more general vegetation category. Discrimination information-based classification results obtained by matching clusters in this scene to labeled clusters for rural Arkansas scene.

TABLE 4-3
Rural Arkansas Classification Results Using Brightness and Greenness Features

Cluster	True classification	Unambiguous labeling	Best match using discrimination information
0	SHA	SHA	**BAR**
1	BAR	BAR	BAR
2	TRE	VEG	**BAR**
3	SPA	VEG	**BAR**
4	TRE	VEG	**BAR**

Note. Unambiguous labeling obtained by collapsing rules for trees and sparse-vegetation into a more general vegetation category. Discrimination information-based classification results obtained by matching clusters in this scene to labeled clusters for suburban Washington scene.

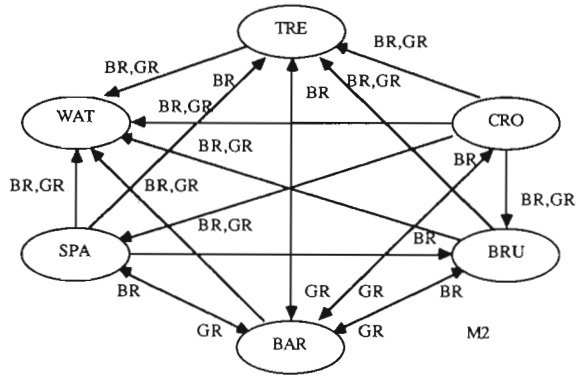


FIG. 4-8. Digraph representation of M2 context which contains BRUsh.

dix. A segmentation based on clustering these images into regions with similar brightness, fractal dimension, and vertical position is shown in the lower left.

A single set of global constraints was used to classify both images:

"Sky is above ground"
"Sky is brighter than trees and shadows"
"Ground is brighter than shadows"
"Trees have a higher fractal dimension (are rougher) than sky, ground, and shadows."

The classification image showing trees in green, shadows in black, sky in blue, ground in red, and areas not classified in white is shown in the lower right. Clusters with a relative frequency of 1% or less were not classified as they tended to be caused by mixed pixels. Tables 4-5 and 4-6 summarize the classification performance for the context-sensitive classifier and the discrimination information-based cluster matcher described earlier. The two scenes are more similar than those examined earlier. As a result the performance of the cluster matcher was more comparable, although still somewhat poorer, than the context-sensitive classifier.

The results in Figs. 4-9 and 4-10 suggest that good global interpretations of images can be obtained using very general scene models. Experiments performed by Navon [20] indicate that visual information appears to be interpreted in a top-down fashion, i.e., global features (objects) are recognized first, followed by a more detailed analysis of local structures. This suggests that the next step might be to embed the context-sensitive classifier in a hierarchical control structure capable of locally refining higher level interpretations. Other areas for improvement include representing more complex topological relationships between clusters in feature space and developing more localized context models to describe constraints between individual instances of objects, e.g., between a particular tree and its cast shadow. A result of using just orderings between cluster means that it is not possible to constrain trees to touch the ground and so we find a few trees floating in the sky in Fig. 4-9. By taking into account the extent (variance) of a cluster, other set relations such as containment, adjacency, and intersection can be represented.

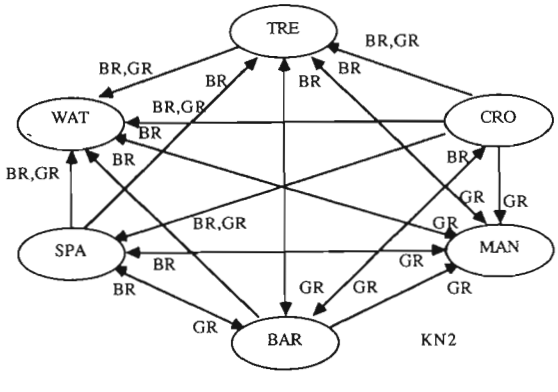


FIG. 4-7. Digraph representation of KN2 context containing CROps and MAN-made materials.

TABLE 4-4
Classification of F2 Scene Using Context Models for KN2, M2, and W1

Cluster	True class	KN2 class	M2 class	W1 class	Majority class
0	SHA/WAT	SHA/WAT	SHA/WAT	SHA/WAT	SHA/WAT
1	BAR	**MAN**	BAR	BAR	BAR
2	TRE	TRE	TRE	**SPA**	TRE
3	SPA	**BAR**	**BAR**	**BAR**	**BAR**
4	TRE	**CRO**	**CRO**	**SPA**	**CRO**

Note. The labelings shown are the most compatible; no unambiguous labelings were obtained.

TABLE 4-5
S1 Classification Results

Cluster number	True classification	Unambiguous labeling	Best match using discrimination information
0	SHA	SHA	SHA
1	SHA	**TRE**	SHA
2	TRE	TRE	TRE
3	TRE	TRE	TRE
4	TRE/GRO	GRO	GRO
5	TRE	TRE	TRE
6			
7	SKY	SKY	**TRE**
8	TRE	**SHA**	TRE
9,10,11			
12	GRO	GRO	GRO
13	SKY	SKY	**TRE**
14	GRO	GRO	**Unknown**
15	GRO	GRO	GRO
16,17,18			

5. SUMMARY

A new approach to the problem of interpreting outdoor scenes has been described. It is based on a context-sensitive classifier which uses relative constraints to describe global relationships between object classes. Contextual models represent the structure of a scene's underlying feature space in terms of stable, physically significant parameters. A discrete relaxation algorithm is used to find unambiguous labelings that satisfy a set of ordering relations between object classes. Unlike rule-based systems, these constraints provide a complete and consistent description of the scene. Scenes that are similar in structure are organized into contexts, each of which is represented by a consistent set of constraints. Instead of attempting to achieve a high degree of specificity and localization within limited domains, the methodology is geared toward recognizing general kinds of objects with little or no human intervention over a wider range of scenes.

TABLE 4-6
S2 Classification Results

Cluster number	True classification	Unambiguous labeling	Best match using discrimination information
0	SHA	SHA	SHA
1	SHA	SHA	SHA
2			
3	SHA	SHA	SHA
4			
5	TRE	TRE	TRE
6	TRE	TRE	TRE
7	TRE	TRE	TRE
8	TRE	TRE	TRE
9,10,11			
12	SHA	SHA	SHA
13,14			
15	TRE	TRE	TRE
16			
17	GRO	GRO	SHA/GRO
18	GRO	GRO	GRO
19	TRE	TRE	TRE
20	TRE	**SKY**	**SKY**
21	TRE/GRO	GRO	**Unknown**
22	GRO	GRO	GRO
23	SKY	SKY	**Unknown**
24,25			
26	TRE	SKY	SKY
27			
28	GRO	GRO	Unknown

Several examples were presented to demonstrate the capability for recognizing general kinds of objects in black-and-white and multispectral imagery acquired by aircraft, satellite, and at ground level. Through a series of experiments, the ability of the system to degrade gracefully in performance when faced with new and unknown situations was also demonstrated.

Three main areas of research remain. The first involves extending the models to allow other kinds of constraints to be represented. Currently, object classes can only be defined in terms of ordering relations between the cluster means. The ability to use information about the extent of an object in the feature space (i.e., its variance) to model topological relations (e.g., containment and adjacency) would provide added flexibility to the representation. A second area of work involves applying the paradigm on a more local basis. This could be investigated by embedding the classifier within a hierarchical control structure, using it to refine an interpretation based on more local context models. Improved segmentation and clustering techniques are a third area to be investigated. The segmentation techniques used in the experiments were relatively simple. As a result, objects were poorly delineated. In order to effectively use geometrical information about the relative size and shape of object regions, the quality of the segmentations needs to be improved. Clustering is currently the most time-consuming part of the process. Finding a more efficient way to extract significant areas that are likely to correspond to the objects of interest is highly desirable.

APPENDIX: GENERATING REGISTERED REPRESENTATIONS OF BRIGHTNESS, TEXTURE, AND SPATIAL INFORMATION

This appendix outlines a method that has been developed for computing registered representations of brightness, texture, and spatial information from a single black-and-white image. The method is based on the concept of averaging within homogeneous regions that have been

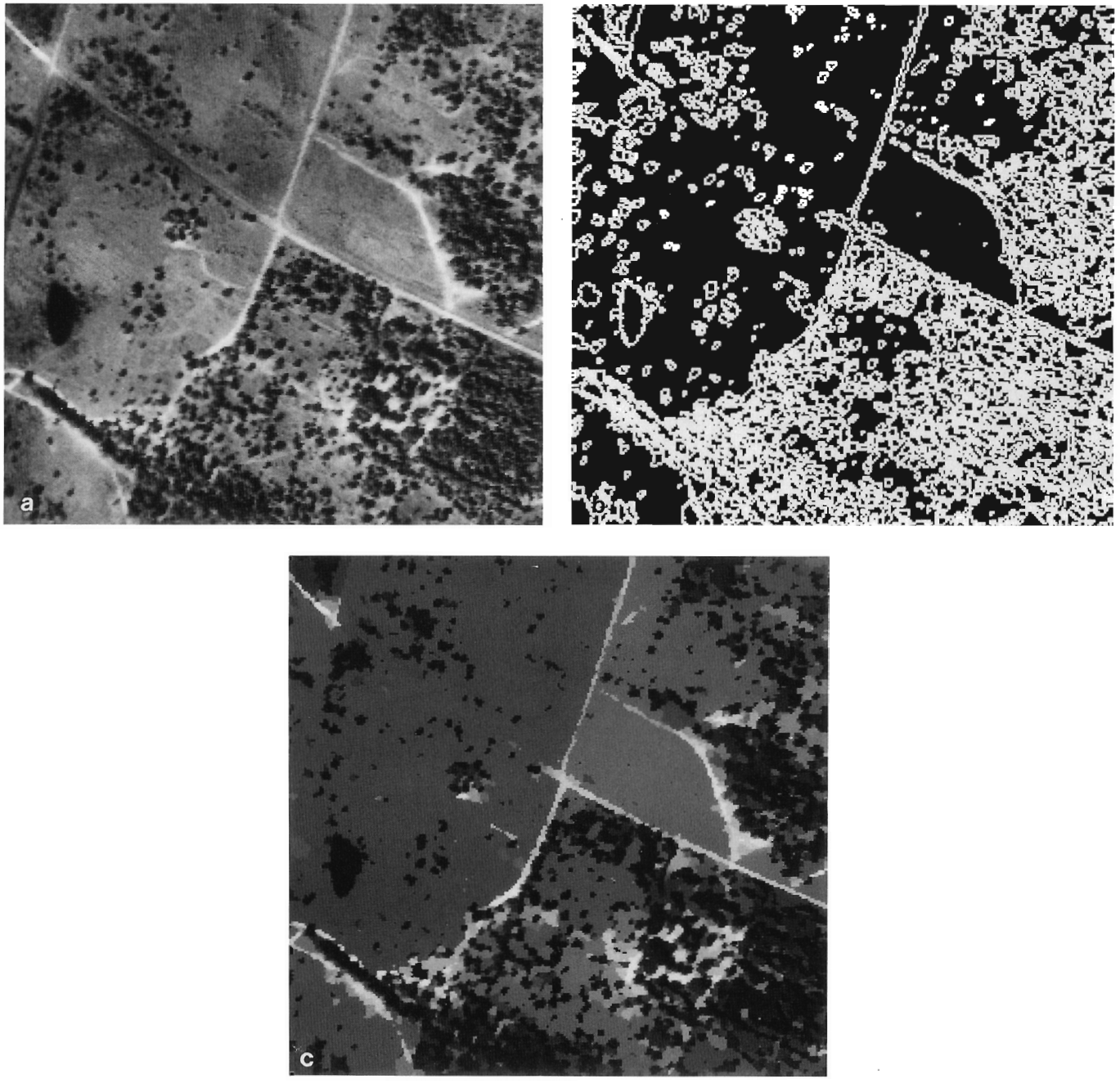


FIG. A-1. Brightness image (a). Boundaries between connected regions (b). Region-averaged brightness image (c).

extracted from a brightness image. The overall process for building registered representations involves:

(1) Segmenting the brightness image into homogeneous connected regions;

(2) Computing the average brightness in each region and spreading the average over all pixels in the region;

(3) Generating one or more images that locally measure physically or perceptually significant textural features such as roughness, coarseness, directionality, etc.;

(4) Computing and spreading average textural features over regions of similar brightness;

(5) Computing and spreading properties (e.g., the position) of each region over itself.

These representations may be accessed either as a registered set of images or as a collection of objects that represent the properties of connected regions. As discussed in Section 4, a clustering algorithm [12] is used to segment the registered set, returning a list of clusters that are

passed on to the classifier. The remainder of this section describes the above process in greater detail.

Brightness segmentation provides a label map of the connected regions for region-averaging. A nonlinear filter is applied to smooth brightness variations within regions but not across edges. Several methods have been developed for this purpose: edge-preserving smoothing [19] anisotropic diffusion [24], and filling-in processes [6]. EPS involves computing, on a pixel-by-pixel basis, the variance over nine neighborhoods about a center pixel, replacing its value with the average over the neighborhood with the lowest variance. After several iterations the original image $I(m, n)$ begins to take on a "cartoon-like" appearance. At this point the filtered image $I'(m, n)$ is segmented into regions of similar brightness. To demonstrate feasibility, a simple region grower was used. The region grower produces a label map $L(m, n)$ which assigns a unique label to each connected region in such a way that $L(m, n) = L(m + m', n + n')$ if $|I'(m, n) - I'(m + m', n + n')| < I_0$. I_0 is a threshold that controls the number of regions that are formed. Since the goal is not to extract potential objects per se, the algorithm is not particularly sensitive to thresholds. A four-connected neighborhood was used and so $|m'| + |n'| = 1$.

The average brightness over each region is computed as follows. Let I_k be the average brightness computed over the kth region R_k,

$$I_k = \sum_{(m,n)\in R_k} I(m, n)/S_k,$$

where S_k is the number of pixels in R_k, and let $I_\mu(m, n)$ be the region-averaged brightness image obtained by spreading the I_k over their corresponding connected regions,

$$I_\mu(m, n) = I_k\delta[L(m, n) - L_k],$$

where δ is the delta function and L_k is the label of the kth region. Figure A-1 shows an aerial panchromatic image (a), the boundaries between connected regions for a threshold $I_0 = 4$ (b), and the region-averaged brightness image (c).

Results from human perceptual studies show the fractal dimension to be highly correlated with subjective measures of surface roughness and to be a useful feature for texture segmentation and discrimination [23]. Other fractal properties that have proven to be useful in describing textures are the lacunarity which indicates the coarseness of the texture [18] and the model-fit error (deviation from fractal behavior) which has been used to discriminate between man-made and natural textures [27]. It has been shown that under certain imaging conditions images of fractional Brownian surfaces are also fractal Brownian [15] and that the fractal dimension is stable with respect to changes in the positions of the illuminant and the viewer [5].

There are many ways of estimating the fractal dimension of images. Here, the fractal dimension is estimated within a sliding rectangular window by the covering method [27]. The covering method involves successive dilations and erosions of the image intensity surface by

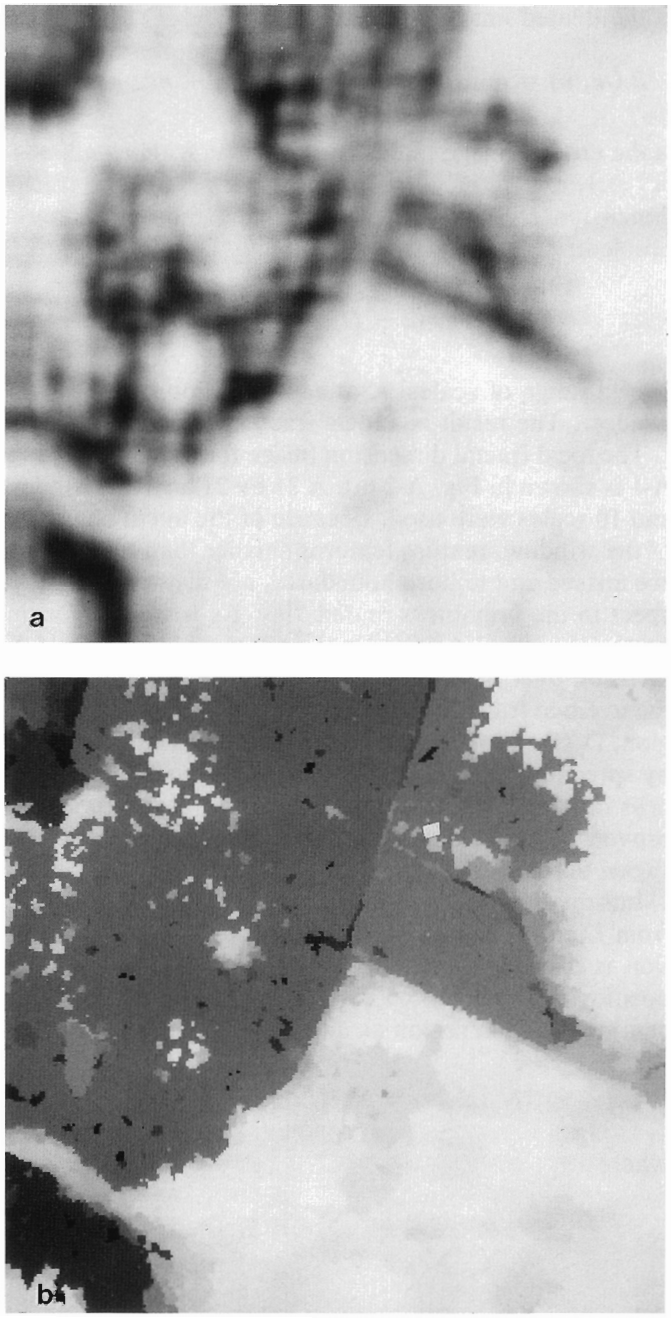


FIG. A-2. Fractal dimension image computed over a 21 by 21 window (a). Result after averaging over connected regions extracted from brightness image (b).

morphological operations. The area of the image intensity surface is estimated over a range of scales as

$$A(r) = (1/2r) \sum_{(m,n)\in W} T_r(m, n) - B_r(m, n),$$

where W is an M by N window;

$$T_r(m, n) = \max\{T_{r-1}(m, n) + 1, T_r(m + m', n + n')\}$$

is the dilated image intensity surface and

$$B_r(m, n) = \min\{B_{r-1}(m, n) - 1, B_r(m + m', n + n')\}$$

is the eroded image intensity surface at scale r, $|m'| \leq 1$, $|n'| \leq 1$, and $T_0(m, n) = B_0(m, n) = I(m, n)$. The fractal dimension D is estimated by solving the least-squares problem:

$$E = \text{Var}[(2 - D)\log r - \log A(r)]$$

over a range of scales, r, at each position of the sliding window. The result is a local fractal dimension image.

The local fractal dimension image for the image in Fig. A-1 is shown in Fig. A-2 (a). A 21 by 21 analysis window and 10 scales were used. Because of the averaging effect of the window, texture features smaller than the window are missed and texture boundaries are displaced with respect to the brightness boundaries. As was done for image brightness, let us average and spread the fractal dimension over the connected regions of $L(m, n)$. If D_k is the average fractal dimension over the kth connected region, $D_\mu(m, n) = D_k\delta[L(m, n) - L_k]$ is the image formed by spreading the D_k over connected regions in the brightness image. The result shown in (b) has significantly improved the separation and sharpened the boundaries between textured regions.

Information about regions can be computed directly from $L(m, n)$. For example, in Section 4.3 vertical position is used to discriminate sky from ground. Vertical position is computed by calculating and spreading the y-centroid of each region as above:

$$N_\mu(m, n) = N_k\delta[L(m, n) - L_k],$$

where

$$N_k = \sum_{(m,n)\in R_k} n/S_k.$$

Other properties of connected regions such as their horizontal position, area, compactness, and elongatedness can be computed in a similar fashion but are not used at present.

REFERENCES

1. R. Bajcsy and L. Lieberman, Computer description of real outdoor scenes, in *Proceedings, 2nd International Joint Conference on Artificial Intelligence, 1974,* pp. 190–194.
2. T. Binford, Survey of model-based image analysis systems, *Internat. J. Robotics Res.* **1,** 1, 1982.
3. M. Carlotto, Histogram analysis using a scale-space approach, *IEEE Trans. Pattern Anal. Machine Intelligence* **9,** 1, Jan. 1987.
4. M. Carlotto, Pattern classification using relative constraints, in *Proceedings Computer Vision and Pattern Recognition, Ann Arbor, MI, June 1988,* pp. 450–456.
5. M. Carlotto and M. Stein, A method for searching for artificial objects on planetary surfaces, *J. Brit. Interplanet. Soc.* **43,** 1990, 209–216.
6. M. Cohen and S. Grossberg, Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance, *Perception Psychophys.* **36,** 1984, 428–456.
7. E. Crist and R. Cicone, A physically-based transformation of thematic mapper data—the TM tasseled cap, *IEEE Trans. Geosci. Remote Sensing* **GE-22,** 3, May 1984.
8. J. Feldman and Y. Yakimovsky, Decision theory and artificial intelligence. I. A semantics-based region analyzer, *Artificial Intelligence* **5,** 1974, 349–371.
9. M. Fischler and R. Elschlager, The representation and matching of pictorial structures, *IEEE Trans. Comput.* **C-22,** 1, Jan. 1973.
10. M. Fischler and T. Strat, Recognizing objects in a natural environment: A contextual vision system (CVS), in *Proceedings, 1989 DARPA Image Understanding Workshop, May 1989.*
11. F. Hall and G. Badhwar, Signature-extendable technology: Global space-based crop recognition, *IEEE Trans. Geosci. Remote Sensing* **GE-25,** 1, Jan. 1987.
12. K. Hartt, M. Carlotto, and M. Brennan, A method for multi-dimensional image segmentation, in *International Geoscience and Remote Sensing Symposium, Vancouver, BC, July 1989.*
13. R. Henderson, Signature extension using the MASC algorithm, *IEEE Trans. Geosci. Electronics* **GE-14,** 1, Jan. 1976.
14. R. Kauth and G. Thomas, The tasselled cap: A graphic description of the spatio-temporal development of agricultural crops as seen by Landsat, in *Proceedings, 3rd Symposium on Machine Processing of Remotely Sensed Data, Purdue Univ., 1976.*
15. P. Kube and A. Pentland, On the imaging of fratal surfaces, *IEEE Trans. Pattern Anal. Machine Intelligence* **10,** 5, Sept. 1988.
16. S. Kullback, *Information Theory and Statistics,* Wiley, New York, 1959.
17. P. Lambeck and D. Rice, Signature extension through the application of cluster matching algorithms to determine appropriate signature transformations, in *Proceedings LARS Symposium, 1976.*
18. B. Mandelbrot, *The Fractal Geometry of Nature,* Freeman, New York, 1983.
19. M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs,* Plenum, New York, 1980.
20. D. Navon, Forest before trees: The precedence of global features in visual perception, *Cognit. Psych.* **9,** 1977, 353–383.
21. Y. Ohta, *A Region-Oriented Image-Analysis System by Computer,* Ph.D. thesis, Kyoto Univ., 1980.
22. C. Parma, A. Hanson, and E. Riseman, *Experiments in Schema-Driven Interpretation of a Natural Scene, COINS Tech. Report 80-10,* Univ. of Massachusetts, Apr. 1980.
23. A. Pentland, Fractal-based description of natural scenes, *IEEE Trans. Pattern Anal. Machine Intelligence* **6,** 6, Nov. 1984.

24. P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Machine Intelligence* **12,** 7, July 1990.
25. A. Rosenfeld, R. Hummel, and S. Zucker, Scene labeling by relaxation operations, *IEEE Trans. SMC* **SMC-6,** 6, June 1976.
26. D. Russell and C. Brown, Representing and using locational constraints in aerial imagery, in *Proceedings, 8th DARPA Image Understanding Workshop, Nov, 1978,* 152–158.
27. M. Stein, Fractal image models and objects detection, in *Proceedings SPIE Conference on Visual Communications and Image Processing,* Vol. 845, Nov. 1987.
28. J. Tenenbaum and H. Barrow, Experiments in interpretation-guided segmentation, *Artificial Intelligence* **8,** 1977, 241–274.
29. L. Valiant, A theory of the learnable, *Comm. ACM* **27,** 11, 1984.
30. S. Zucker, A. Rosenfeld, and L. Davis, General purpose models: Expectations about the unexpected in *Proceedings, 4th International Joint Conference on Artificial Intelligence, 1975,* pp. 716–721.