Computational Image Interpretation Models: An Overview and a Perspective

Demetre P. Argialas
Department of Civil Engineering and Remote Sensing and Image Processing Laboratory, Louisiana State University, Baton Rouge, LA 70803
Charles A. Harlow
Department of Electrical Engineering and Remote Sensing and Image Processing Laboratory, Louisiana State University, Baton Rouge, LA 70803

ABSTRACT: Methodologies of pattern recognition, image analysis, computer vision, and knowledge-based expert systems are reviewed. The potential of their integration for effective interpretation of remotely sensed images is delineated. The advantages and limitations along with the basic similarities and differences between the approaches are examined. The evaluation takes place so as to indicate the gradual shift from tone/color to site/association descriptions, from pixel- to object-based representations, from spectral to spatial and structural methods, from numeric to symbolic computations, from procedural to declarative programming, and from domain-independent to knowledge-based systems. A number of knowledge-based image interpretation systems are examined. The complexity of aerial and high resolution satellite images emphasizes the need for exploitation of the knowledge employed by expert photointerpreters and its encapsulation in expert consultation systems. Explicitly coded knowledge in such systems can be replicated, taught, criticized, and expanded and it offers the promise for smarter image interpretation systems. It is necessary to develop methodologies defining how photointerpreters assimilate facts, articulate descriptions, and perform their reasoning.

INTRODUCTION

STATISTICAL PATTERN RECOGNITION has been employed for image classification from remotely sensed images for the past 20 years. Owing to its assumptions, it is ideally suited for applications in which the pattern classes can be described by a set of numerical measurements (spectral signatures), which can be represented in a vector form (Swain and Davis, 1978; Schowengerdt, 1983). Image processing has been employed mainly for image enhancement applications. The results of statistical pattern recognition and image processing techniques have been rather crude compared with those of a skilled photointerpreter (Philipson, 1980; 1986). Image interpretation is more than reading pixels from an image as it involves seeing and understanding, which requires both the identification of image pattern elements (i.e., tone, color, size, texture, shape, pattern, height, shadow, site, association) and the analysis and articulation of conceptual knowledge, based on diverse stereotyped models and heuristic rules employed by expert interpreters.

Artificial intelligence (Winston, 1984) and knowledge-based expert systems (Harmon and King, 1985) have provided powerful methodologies for the development of computer programs that have the potential to represent expertise in image interpretation. A major area of contribution of the artificial intelligence and expert system paradigm to pattern analysis has been the study of how domain specific and heuristic knowledge can be represented and used to control the process of extracting meaningful descriptors and objects from images. In image analysis, as in artificial intelligence, a paradigm shift has emerged from domain independent to knowledge-based representation techniques (Figure 1). The flexibility, power, and effectiveness of combining artificial intelligence, pattern recognition, and image analysis techniques have been demonstrated with the research prototypes already built.

Previous reviews of pattern analysis and expert system methodologies for image interpretation have been contributed by Estes *et al.* (1983), Mooneyhan (1983), Campbell and Roelofs (1984), Shapiro (1985), Fabbri *et al.* (1986), Rosenfeld (1986), Tailor *et al.* (1986), and Friedl *et al.* (1988). The comparison and contrasting of human and computer assisted approaches to image interpretation have been described by Estes *et al.* (1986). Parallel to the fusion of artificial intelligence and image analysis techniques, there is a trend for knowledge-based geographic information systems (GIS), in which techniques from artificial intelligence are merged with GIS techniques for new designs and products. Robinson *et al.* (1986, 1987), Fisher and Mackaness (1987), and others have reviewed developments in this field.

PATTERN RECOGNITION AND CLASSIFICATION

Pattern recognition is defined as the categorization of data into identifiable classes through the extraction of significant features or attributes of the data (Tou and Gonzalez, 1974). Basic functions required to recognize objects in images are preprocessing, feature selection and detection, segmentation, description, recognition, and classification. Feature detection, segmentation, and description are explored predominately in image analysis and computer vision, whereas pattern recognition emphasizes feature selection and classification techniques (Fu and Rosenfeld, 1976). Description and segmentation are essential components of the structural and syntactic pattern recogntion approaches (Gonzalez and Thomason, 1978).

The methodology of pattern recognition applied to a particular problem depends on the data, the models about the data, and the information that one is expecting to find within the data (Bezdek, 1981). The data may be qualitative, quantitative, numerical, pictorial, textual, linguistic, or any combination of the above. Pictorial data carry information about the objects in the scene depicted in the image (Figure 2). The manner in which this information can be described and organized so that relationships between the scene objects can be identified is ascribed to the structure of the image (Bezdek, 1981). Image information can be described at many levels of abstraction. A description may range from one in terms of meaningful attributes of the scene depicted in the image, to one that describes only the spatial variation of intensity. Any of these descriptions can be expressed with a model that captures only the relevant features of the image in that level of abstraction and leaves others unspecified. The role of a model is to convert information in the image into usable forms and, therefore, enable the inference of

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 56, No. 6, June 1990, pp. 871–886.

0099-1112/90/5606–871$02.25/0

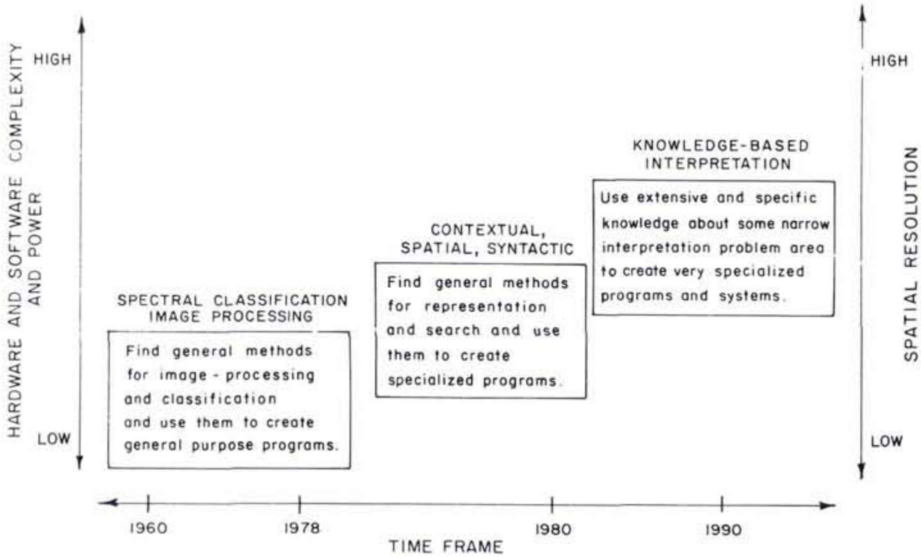

FIG. 1. The evolution and shifting of image interpretation techniques.

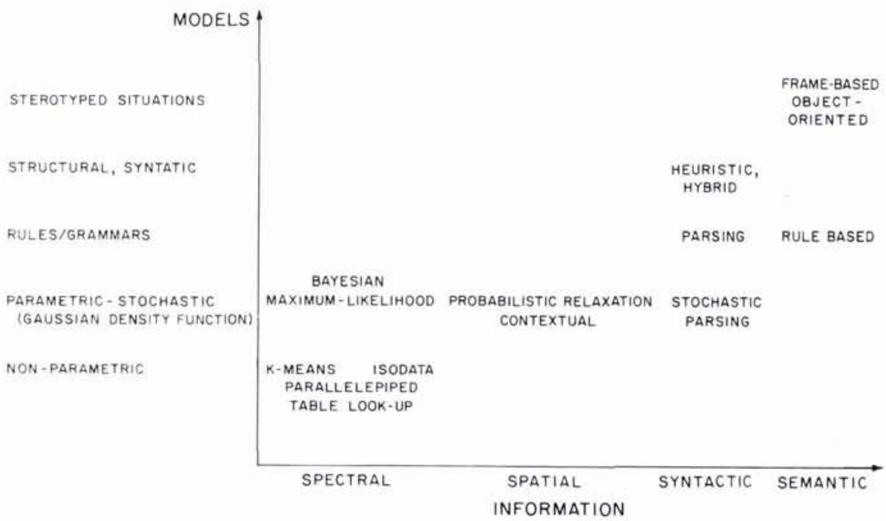

FIG. 2. Dependence of pattern analysis techniques on models and information.

objective properties of the objects being studied. To do so, the model must assimilate the data and make them compatible with the search and matching strategies to be used. Each search and matching strategy corresponds to a different pattern recognition methodology. This is the reason for the diverse approaches to pattern recognition (Figure 2), e.g., mathematical or statistical, grammatical, syntactic or structural, and heuristic or descriptive (Tou and Gonzalez, 1974; Mero and Vamos, 1981; Nevatia, 1982).

STATISTICAL AND CONTEXTUAL MODELS

The decision theoretic or statistical pattern recognition approach is ideally suited for image interpretation applications where the pattern classes can be described by a set of numerical measurements or features (Figure 3). The measurement vector is composed of the gray level or spectral properties of the classes considered. The power of this approach is dependent on (1) the availability of features that are invariant to the expected changes within the pattern classes, (2) the amount of discriminating information contained in the measurements or features, and (3) the effective utilization of this information in a suitable classification algorithm.

The first step in this approach is the selection and extraction of a set of measurements or features from the pattern classes. The choice of features is problem dependent. In processing Landsat and SPOT satellite data or airborne multispectral data, one can use all the available bands or any subset thereof. To reduce the dimensionality of the data, and to eliminate highly correlated spectal bands, principal component analysis (Karhunen-Loeve technique) or canonical analysis are often used for feature (band) selection. These techniques can also be used when the standard band vector is augmented by measures of texture, elevation, or multiple image sets (Swain and Davis, 1978; Moik, 1980; Schowengerdt, 1983; Jensen, 1986).

The classification of the feature vectors usually takes place by the use of a similarity measure, such as a distance measure, a discriminate function, or a likelihood function (Figure 3). If a complete set of discriminatory features for each pattern class can be determined from the data, then the patterns are

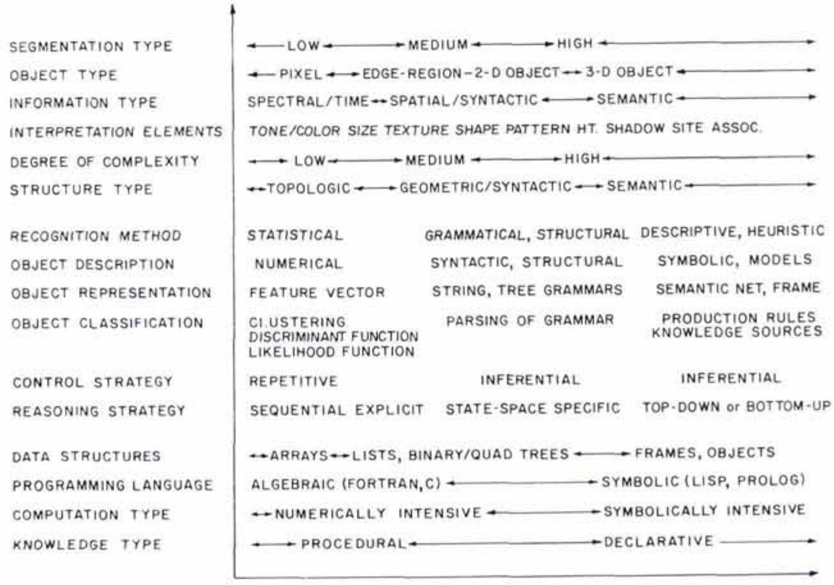

FIG. 3. Computational image interpretation models and their association to key concepts.

represented by a "feature vector," and the recognition and classification of patterns may be reduced to a simple matching process or a "table look-up" scheme. To take noise and distortions into consideration, as well as overlapping classes, statistical methods have been employed. In such methods, each pattern class is represented by a class-conditional probability density function, and the classification of unknown patterns is based on a parametric or nonparametric statistical decision rule. To determine the parameter values of a decision rule or the parameter values and the form of the class-conditional probability density function, various supervised and unsupervised learning algorithms have been suggested (Swain and Davis, 1978; Moik, 1980; Schowengerdt, 1983; Jensen, 1986). K-Means and ISODATA are methods of unsupervised clustering that determine, on the basis of observed measurements, the pattern classes responsible for generating such measurements (Niblack, 1986).

Statistical pattern recognition techniques have been employed for image classification in remote sensing for the past 20 years for such diverse applications as land-cover identification, tree species classification, water pollution mapping, geologic classification, and thermal mapping (Lillesand and Kiefer, 1987). These methods, however, have been based primarily on the multispectral characteristics of individual pixels without considering spatial context, that is, relations among neighboring points. Thus, they result in characterization of spectral classes rather than object identification and description, which are at the core of standard photointerpretation techniques. For example, while individual pixels can be classified as water bodies, it is not possible to infer that a certain patch of pixels comprise a lake or river. This shortcoming is partially due to the lack of utilization of knowledge pertaining to the scene objects during the recognition and classification processes.

As the spatial resolution of the acquired remotely sensed data increases (MSS, TM, SPOT), spatial context can contribute to the interpretation of complex objects (Figure 1). Innovative models have therefore been developed to take into account temporal, spatial, and contextual information in addition to spectral information (Figure 2). Temporal information was employed through signature extension methods that employ the change of spectral features over time. However, registration and calibration problems seriously hamper the application of these methods. Classification of multispectral data by extraction and classification of homogenous objects has been carried out by Kettig and Landgrebe (1976). Spatial logic techniques have been employed involving spectral stratification, region formation, and iterative classification algorithms (Merchant, 1984). Classification accuracies have been improved by employing ancillary map data such as slope, aspect, and elevation (a) before classification for stratification, (b) during classification by modifying the *a priori* probabilities in the maximum likelihood classifier or as another layer in the classification process, and (c) for post-classification sorting by resolving problematic spectral classes (Hutchinson, 1982; Richards *et al.*, 1982; Schowengerdt, 1983; Jensen, 1986; Niblack, 1986). However, even these spatial/contextual classification methods did not address structural and semantic relations between features or objects, that is, the underlying structure of the scene represented in the image. For recognition of complex and structured image patterns, such as highways and airports, drainage patterns, and landforms, the "feature vector" representation and the statistical classification approaches are not adequate.

STRUCTURAL MODELS

The weakness of the statistical approach in classifying complex image patterns, as these usually appear on high resolution aerial images, is its inability to cope with what is thought of intuitively as the "structure" of the pattern. Structure has been defined as the configuration of elements, parts, or constituents in a complex entity or the interrelation of parts or the principle of organization in a complex entity (Morris, 1979). There does not seem, though, to be anything in nature of reality that is structured *a priori* (Robinson and Petchenik, 1976). Structure is what is attributed by an expert to an arrangement of components which has more meaning than that obtained from simply an aggregation of the parts. While an image pattern may have the potential for being regarded as having structure, the actual conception of that structure must be provided by an expert.

A structural description of a pattern can be described as an organization of subpatterns, objects, or elements. A subpattern can be again considered as an organization of further subpatterns or objects, and so on. An organization is a complex of relationships that subsist between the elements that are recognized. Thus, a method of recognizing a pattern is to search and find if a particular organization exists, that is, search for a certain configuration of objects satisfying particular relationships (Narasimhan, 1974). Structural descriptions are often employed by expert photointerpreters in identifying complex image patterns by articulating their several features. For example, an interpreter can identify a certain drainage pattern as being of a certain type by recognizing that (1) it is composed of a main stream, from which issue primary tributaries, and from which main stream and primary tributaries issue secondary tributaries, and (2) it is characterized by specific attributes of and relationships among the main stream, and the primary and secondary tributaries (Argialas *et al.*, 1988). Therefore, to recognize a drainage pattern, it is necessary to seek, describe, and represent syntactic and semantic components of the parts and the relations among them. Classification techniques based on such structural descriptions of image patterns have been proven to be more comprehensive schemes than classification techniques based on a per-pixel feature vector, which is a degenerate description, with the structural information not included.

To effectively describe and represent the photointerpreter's knowledge, expressing the structure of complex patterns, that knowledge must somehow be stored and made available during the analysis and recognition processes. Clearly, the feature vector approach is not adequate for such a representation. It has been suggested that such knowledge can be recorded as a structural model, which can represent images in terms of perceived image structure (Pavlidis, 1977). A structural model can be used to guide the choice of image or object partitioning, attribute extraction, and classification operations. Structural models are based on defining primitive pattern elements and identifying allowable structures in terms of relationships among primitives and substructures that combine primitives (Kanal, 1974). Structural models combine domain specific knowledge about the scene context together with specific symbolic representation in order to produce a strutural description (Duda and Hart, 1973; Rosenfeld and Kak, 1982). Any model that is at least incidentally concerned with decomposing a pattern into subpatterns (alternatively, synthesizing it from subpatterns) is called a linguistic (Rosenfeld and Kak, 1982), structural (Pavlidis, 1977), or syntactic (Fu, 1974) model (Figures 2 and 3). A model that utilizes in any way the syntactic description of a pattern is called a syntax-directed, descriptive, implicit, or relational model. If the grammar on which such a description is based is explicitly directing the analysis, then the model is called syntax-controlled, grammatical, or explicit (Fu and Swain, 1971).

Grammatical or Linguistic Models

In the grammatical or linguistic approach, patterns are decomposed in a hierarchical concatenation of subpatterns analogous to the syntactic structure of languages, and are defined by a formal grammar. The essence of this approach lies in the selection of pattern primitives and subpatterns, the assembling of the primitives and their relationships into pattern grammars, and analysis and recognition in terms of these grammars (Figure 3). Grammatical pattern recognition has adapted the techniques of formal language theory, which provide both a notation (grammar) and an analysis mechanism (parsing), to the problem of representing and analyzing patterns containing a significant syntactic content (Fu and Swain, 1971; Fu, 1974; Pavlidis, 1977; Gonzalez and Thomason, 1978; Fu, 1980a; Pavlidis, 1982).

The syntax of an image pattern is defined as the juxtaposition and concatenation of subparts. The rules governing these compositions are often specified by the "pattern grammar." A pattern grammar is composed of a set of rules of syntax which define the permissible or desirable relations between these subparts. A syntax-controlled analysis employs the syntax of the grammar in the analysis process. The language which provides the structural description of patterns, in terms of a set of pattern subparts and primitives and their composition relations, is often called "pattern description language" (Fu and Rosenfeld, 1976). A pattern grammar characterizes a structural pattern in the syntactic approach, in a similar way that a measurement vector (spectral signature) characterizes a pattern class in the statistical approach (Figure 3).

A pattern primitive is defined by two components: a token or symbol from a finite alphabet, and an associated list of attributes consisting of logical, numerical, or vector values (Thomason and Gonzalez, 1981). Commonly used primitives include edges, lines, curves, and angles (You and Fu, 1979; Fu, 1982a). Primitives are described by topologic and geometric attributes such as length, position, and orientation. Because the primitives do not embrace any structural information, they are usually derived by non-syntactic methods (Fu, 1974). The feature extraction problem in the statistical pattern recognition approach and the primitive extraction problem in the syntactic approach are similar in nature except that the primitives in the syntactic approach represent subpatterns whereas the features in the statistical approach may be any set of numerical measurements taken from the pattern (Fu, 1974). The primitive extraction problem is equivalent to the low-level segmentation methods (Marr, 1982) discussed later.

Each pattern class is usually encoded as a string, tree, or graph, composed of primitives from the set of permissible primitives of the language. Then a pattern grammar is constructed with the property that the language it generates consists of sentences or patterns which belong exclusively to one of the pattern classes. Subsequently, this grammar is used for pattern classification by employing a procedure for determining whether or not a given pattern of unknown origin represents a valid statement of the language generated by that grammar. The procedure used to determine whether or not a string, tree, or graph represents a sentence which is grammatically correct with respect to a given language is called parsing and is performed by the parser or analyzer (Figure 3). One way that the parser may categorize the input patterns is by matching the string, tree, or graph of an unknown pattern against the diverse prototype or candidate strings, trees, or graphs (Fu, 1976; Fu, 1980a).

To provide a way to deal with geometrical aspects of grammatical inference, syntactic-semantic models were developed, through attributed grammars (Tsai and Fu, 1980; Fu, 1982a). In attributed grammars, each primitive has associated with it geometric attributes such as length, chord length, curvature, and symmetry, and the grammatical rules have a syntactic and a semantic component (Fu, 1982a). Learning from sample patterns, that is, learning a grammar from a set of sample patterns, is still in its early development in relation to learning capabilities which would be acceptable as a general tool for syntactic pattern recognition (Fu, 1980a). Conventional parsing requires an exact match between the representation of an unknown pattern and those generated by the pattern grammars, and thus it does not apply to noisy and distorted patterns. Stochastic grammars have been developed to take into account measurement noise, distortions, and ambiguity due to lack of complete knowledge about the characteristics of pattern classes (Fu, 1977; Gonzalez and Wintz, 1977). To circumvent exact matching, a parsing procedure, called error-correcting parsing, was developed involving a selected similarity measure (distance measure or likelihood function) (Tsai and Fu, 1980; Fu, 1981).

The grammatical approach has been viewed as two distinct processes – primitive extraction followed by grammatical description. In separating the grammatical analysis of structure from the extraction of primitives, each process is excluded from information available to the other (Watanabe, 1971; Kanal and Chandrasekaran, 1972). This is a weakness similar to the one encountered in the bottom-up segmentation methods, discussed later, where the extraction of low-level objects proceeds in ignorance of the *a priori* restrictions known to the high-level models.

Parsing becomes available for grammatical analysis only if the class of patterns of interest can be specified by a formal grammar. This has been achieved with some success for restricted applications (Fu, 1977; Fu, 1980b; Kanal and Rosenfeld, 1981; Fu, 1982c). Li and Fu (1976), Fu (1976), and Brayer *et al.* (1977) described a tree grammar approach which interprets highways and rivers in Landsat images.

HEURISTIC OR DESCRIPTIVE MODELS

The search for grammars describing large classes of natural patterns has not yet been very successful (Nevatia, 1982). When a formal grammatical model is not explicitly used, the terms "ad-hoc" or "heuristic" or "descriptive" (Figure 2) are used to describe the image model (Kanal, 1974). Grammars can be employed in a heuristic capacity for generating information structures more than for generating the actual data structures. For example, labeled graphs have been used to represent information structures, so that each node of the labeled graph corresponds to an object, and is labeled with a list of property names and property values that hold for that object; and each arc between pairs of objects is labeled with a list of names and values of the relations that hold between that pair of objects (Findler, 1979; Rosenfeld and Kak, 1982; Fu, 1981; Shapiro, 1980; Thomason and Gonzalez, 1981; Winston, 1984). Classification is often implemented by a series of tests that are designed to evaluate the occurrence or nonoccurrence of certain subpatterns or primitives, or certain attributes and their actual values, or certain combinations of them. These tests may be embedded within a decision tree, a rule-based system, or a hierarchical classification model.

Argialas *et al.* (1988) designed and implemented structural models that expressed drainage patterns in terms of their constituent elements, their attributes, and the relationships among them. Eight pattern types were quantitatively described and classified. The methodology to classify these patterns consisted of drainage pattern models, hierarchical and relational models, attribute extraction, and classification strategies.

HYBRID MODELS

An efficient pattern recognition system for aerial images requires integration of statistical, linguistic, and heuristic tools in various stages of the design. The term "hybrid system" has been employed to describe pattern recognition systems in which both numerical (statistical, decision theoretical) and qualitative (structural, syntactic, heuristic) techniques are combined (Nadler, 1982). Fu (1982b) identified five ways to mix the decision-theoretic approach and the syntactic approach. They are: decision theoretic followed by syntactic approach, use of stochastic languages, stochastic error-correcting syntax analysis, matching of stochastic graphs, and use of stochastic attributed grammars. In the decision-theoretic followed by syntactic approach, pattern primitives are recognized by a decision-theoretic method (Fu, 1982b). For example, each pixel in a Landsat image can be classified by a decision-theoretic method, such as the maximum-likelihood classification rule using multispectral measurements. Structural (or spatial) relations among various pixels can be described by a syntactic method. Specifically, the structure of rivers was represented by trees with water-like pixels, and was characterized by a tree grammar. Consequently, the recognition of rivers from all water-like pixels was accomplished by using a tree automaton (Fu, 1981; Fu, 1982b). Another example of mixing the two approaches has been the explicit inclusion of semantic evaluations simultaneously with syntactic analysis by means of attributed grammars (You and Fu, 1979; Tsai and Fu, 1980).

IMAGE ANALYSIS AND COMPUTER VISION

The domain of image analysis and computer vision extends beyond that of image classification, which is the core subject of pattern recognition techniques. It encompasses extraction of objects, description of their structure and mutual relationships, and analysis of their similarities and differences with other known objects (Nevatia, 1982; Suen and Mori,1982). Image analysis includes image coding, compression, enhancement, restoration, reconstruction, description, representation, segmentation, and recognition techniques. The first five techniques have been developed since the late sixties and have traditionally been the concern of the field of image processing, while the last four have been recently evolved, primarily in association with computer vision and artificial intelligence techniques.

Computer vision is the study of computational systems that interpret natural scenes, that is, they produce descriptions of a scene from digital images of that scene. The emphasis is in the design and implementation of effective methods that represent and exploit knowledge of the world and knowledge of how images are formed. This knowledge includes properties of sensors, geometry and irradiance relations of stereoscopic models (Alvertos *et al.*, 1989), laws of physical optics, and information about possible configurations in the world. Computer vision grew out of image analysis, pattern recognition, and perceptual psychology. Computer vision is often considered part of the field of artificial intelligence and knowledge-based systems because it aims at building machines that behave intelligently in perceptual domains (Ballard and Brown, 1982). Artificial intelligence is the study of activities that require intelligence (Winston, 1984). Knowledge-based expert systems (KBES) is a field of artificial intelligence that emphasizes specific but difficult problem solving requiring domain specific knowledge (Hayes-Roth *et al.*, 1983; Harmon and King, 1985).

In the earlier developments in image analysis, there was no sharp dividing line between knowledge-based (high-level, semantic) models and general purpose (low-level,non-semantic) models (Zucker *et al.*, 1975), partially because the knowledge-based methodology was not available to provide the tools to do so (Figure 1). Vision systems were programmed in procedural languages, which did not provide for separation of declarative domain knowledge from search and control algorithms (Figure 3). Procedural representations of knowledge did not provide a means for representing expertise in an intelligible, inspectable, and explicit form. With the growing need for incorporating domain-specific knowledge in image analysis, and the subsequent development of knowledge-based techniques, the distinction between the non-semantic and semantic information became feasible (Marr, 1982). A major area of contribution of the artificial intelligence and expert system methodologies in computer vision has been the study of how domain specific knowledge can be represented and used to control the process of extracting meaningful descriptors and objects from images (Horn, 1978; Marr, 1982). It is now possible to use hybrid programming environments, where object-oriented programming, frames, rule-based programming and methods, or active values can be employed at the appropriate level of problem representation. As a result, image analysis, computer vision, and artificial intelligence are inevitably and inextricably linked.

IMAGE DESCRIPTION, REPRESENTATION, AND SEGMENTATION

From a conceptual point of view, information in images can be described at many levels of abstractions or domains. An abstraction may represent the scene depicted in the image, the three-dimensional (3-D) objects of the scene, or the image domain (Clowes, 1971). The scene domain consists of knowledge and models of the semantic objects depicted in the scene. An object can be any relevant feature in the scene which is to be detected and recognized. The world domain consists of the representation of physical objects in three-dimensional space. The image domain describes only the spatial variation of intensity obtained through the imaging geometry consisting of the source, object, and sensor. The image domain is a domain of observable facts which can be obtained from the digital image while the scene domain is an abstract domain in which objects are modeled and relations are defined. The world domain consists of all physical objects with their surfaces and surface material which relate to observable properties such as reflectance and texture. Each object, according to the imaging geometry, will project an image in the image domain. The image domain is composed of pixels, which are single points, and patches, which are composed of connected sets of pixels with some uniformity property such as uniform gray level or texture. In general, patches will have to be grouped together to indicate a surface of a physical object in the world domain. A part of the image, perhaps a set of patches which corresponds to an object surface, is often called a region.

Another conceptual classification recognizes three levels of representation: high, medium, and low (Ahuja and Schachter, 1983). High-level representations involve highly semantic (meaningful) objects relating to image understanding itself. High-level representations allow the interpretation of an image in terms of the goals of the analysis. Although high-level objects are the ultimate concern, even these objects must be defined through medium-level objects and primitives. Medium-level representations provide simple aggregations of the basic primitives. Primitives are the most basic image elements that do not contain any semantic information and that can be meaningfully and adequately incorporated in the higher level of representations (Marr, 1982).

Most approaches for obtaining an image description start with the primitives, and synthesize higher representations by aggregating lower-level objects to higher-level objects in bottom-up hierarchical procedure. The first step, in this procedure, is the identification and labeling of primitives. Further grouping of these primitives may take place, to identify higher-level, more meaningful objects. This process of partitioning or synthesizing an image into its constituent objects is called segmentation (Thomason and Gonzalez, 1981; Rosenfeld and Kak, 1982). Segmentation methods can be characterized according to the level of semantics employed into three levels: low, medium, and high (Figure 3). A low-level segmentation method has as input the grey level image pixels and as output the primitives such as edges, blobs, lines, and arcs (Marr, 1982). Low-level segmentation does not depend on one's knowledge or expectations about the particular situation, but on what it is possible to compute from an image. Medium-level segmentation corresponds to the intermediate details of the analysis — the nodes which set up partial results from the primitive elements or which decompose the highest level of description into substructures (Mero and Vamos, 1981). A segmentation method is considered high level if its output is the final interpretation of the image. High-level segmentation approaches are more sensitive to the specific needs of recognition and to the context of the image (Ballard and Brown, 1982; Marr, 1982).

LOW- AND MEDIUM-LEVEL SEGMENTATION METHODS

Virtually hundreds of low- and medium-level image analysis algorithms have been proposed in the literature over the past 20 years (Gonzalez and Wintz, 1977; Ballard and Brown, 1982; Pavlidis, 1982; Rosenfeld and Kak, 1982; Haralick, 1985). Not surprisingly, this is still an active area of research because of its importance as a major processing step in any practical application. The choice of one segmentation technique over another is dictated mostly by the characteristics of the problem being considered (Figure 4). Data structures for representations of low- and medium-level objects have been described by Shapiro (1979) and Samet (1980).

The most common operators for low-level segmentation are neighborhood operators which examine the value of a small neighborhood of pixels around a given pixel and produce a resultant value that is a function of all pixel values in the neighborhood. Low-level neighborhood operators have focused upon edge detection and strategies for connecting edge points into lines. This approach has intuitive appeal because the eye is sensitive to edges and humans can recognize objects from the object's outline (Marr, 1982). These approaches have encountered difficulty in practice because global parameters may dictate correct interpretations. Reviews of edge operators and attempts to overcome some of these problems are given in Brooks (1978) and Peli and Malah (1982).

Medium-level segmentation algorithms include region growing, histogram thresholding, boundary detection, multispectral classification, probabilistic relaxation, and texture and shape analysis (Figure 4). Region growing methods were developed to overcome some of the problems with edge detection (Zucker, 1976). In region growing one attempts to locate areas of the image that share some uniformity property, such as uniform density, similarity proximity, and good continuity and closure. The most common property used by far is uniformity of gray level. The hope is that it is easier to locate regions with a uniform property than to measure dissimilarity with edge detectors. Pavlidis (1977) described split and merge techniques for region formation using various methods to measure region similarity. A tool often used with split and merge techniques is the region adjacency graph. It is accessible by region label and contains, for each region, a list of all adjacent region labels. All the various thresolding methods using histograms are a form of region formation (Ohlander *et al.*, 1978; Trivedi and Harlow, 1985; Kapur *et al.*, 1985).

Color (Ohlander *et al.*, 1978), multispectral (Kettig and Landgrebe, 1976), and texture properties (Chen and Pavlidis, 1979) have been used to some extent in region formation and object recognition. For multispectral classification, unsupervised clustering techniques are used to partition the measurement space into spectrally distinct classes, followed by image labeling through reference to training data. Hybrid supervised and unsupervised approaches are also available (Schowndegerdt, 1983). The use of textural representations to supplement the spectral information of remotely sensed images generally produces improved accuracies (Haralick, 1979; Conners *et al.*, 1984; Harlow *et al.*, 1985) Relaxation techniques have been employed to reduce ambiguities in segmentation by taking into account local context. Prior likelihoods are usually assigned to each pixel as belonging to each of a set of image objects. Further refinement of these probabilities takes place iteratively according to probabilities of the neighborhood of each pixel. Probabilistic relaxation labeling techniques have been applied for the enhancement of rivers in Landsat images (Rosenfeld *et al.*, 1976; Zucker, 1976).

Shape — the form of an object — is one of the most important interpretation elements. Many natural and manmade interpretative features such as deltas, oxbow lakes, and drainage patterns are named from distinctive shapes and forms. Shape analysis consists of producing a description of the form of an object that can be used for identification, grouping, or further processing of the object (Shapiro, 1980). Shape analysis techniques can be divided into structural and nonstructural techniques. Nonstructural techniques describe a shape as a vector of scaler features. For example, the boundary of the shape can be expressed as a function — the function approximated by a Fourier series — and the coefficients of the Fourier series used for shape descriptors (Gonzalez and Wintz, 1977). A structural

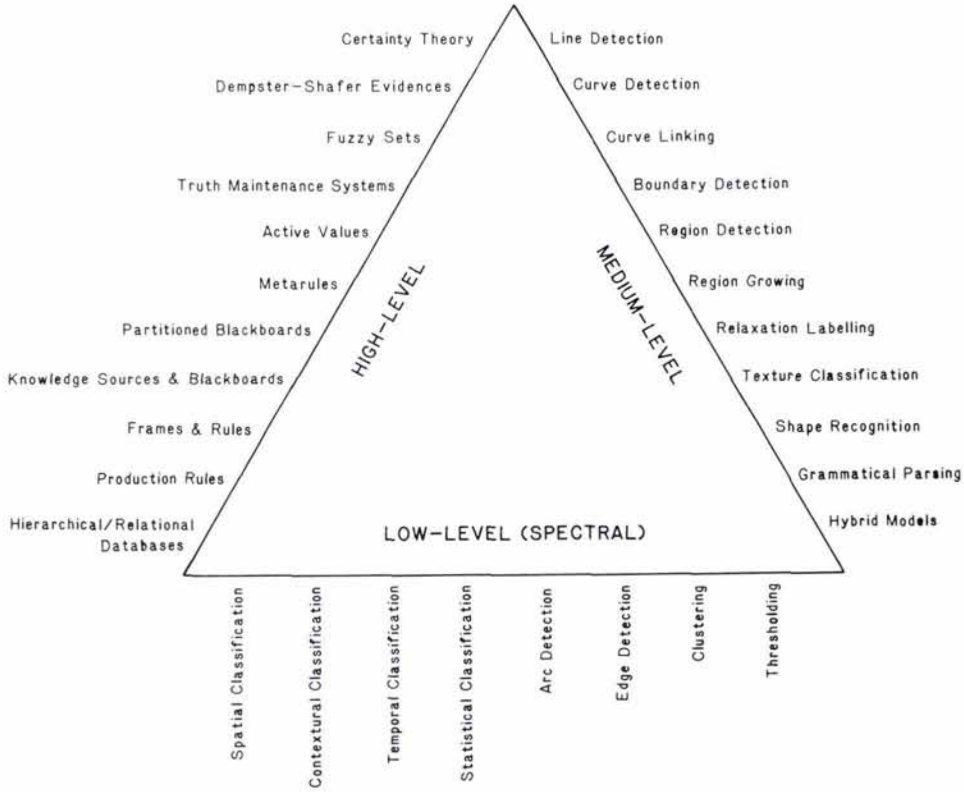

FIG. 4. Potential combinations of image interpretation techniques.

description of the shape of an object consists of primitives, subparts, properties of the subparts and of the whole, and relationships among the subparts. Structural descriptions of shape, like those described earlier under grammatical models, tend to be more robust than nonstructural descriptions and can describe more complex objects (Shapiro, 1980).

HIGH-LEVEL (KNOWLEDGE-BASED) SEGMENTATION METHODS

High-level segmentation utilizes whatever domain-specific knowledge is available about the class of scenes that it is to "understand." Domain-specific knowledge has many forms, including descriptive definitions of entities, concepts, and objects and their relationship to each other and criteria for making decisions. With the fusion of artificial intelligence techniques in computer vision, knowledge representation started to shift from procedural to declarative forms. A procedural representation of a fact is a set of instructions that, when carried out, arrive at a result consistent with the fact, while a declarative representation of a fact is an assertion that the fact is true. Usually, in the procedural instructions, there is no explicit statement of a fact. The fact is contained only in the list of results of the procedure. Several declarative and hybrid knowledge representation methods, such as semantic nets, predicate logic, rules, frames, and blackboards, have been employed for processing semantic information in computer vision (Figure 3). Symbolic problem solving methods, such as modus ponens, resolution, and inexact reasoning, have been employed for drawing inferences (Ballard and Brow, 1982). Control schemes such as forward and backward chaining, depth first, and breadth first search are employed for controlling attention during problem solving (Hayes-Roth *et al.*, 1983).

SEMANTIC NETWORKS

A semantic network can be used to represent declarative knowledge. An entity (node) of the network represents general objects, and arcs represent relationships between objects (Ballard and Brown, 1982). A token or instance of a node is a specific example of the general type. This is described by an IS-A link or arc. For example, a road might be a node and a road in a specific location would be an instance of the object (Harlow *et al.*, 1986). This results in a copy being made of the node which describes this specific example. This might include specific values for the properties of the node, for example, the width of the road. Another special arc or link is the a-kind-of or AKO link, which indicates generalization (Winston, 1984). For example, a car is a-kind-of vehicle. Properties in a semantic net are inherited through the AKO link. Semantic networks are static representations and require an algorithm to operate on the network to generate and evaluate inferences.

PREDICATE LOGIC

Knowledge representation in first-order predicate logic provided the basis for logic programming languages such as PROLOG. The formalism permits expression of facts and rules. Simple declarative facts can be expressed as instantiated predicates (relations) and rules as "predicate X if predicates A, B, C." Reasoning is implemented as depth-first search and backward chaining (backtracking). Inferences (deductions) are made by the resolution principle. Predicate calculus offers consistency, formality, and expressiveness. However, no hierarchical framework exists in which rules can be embedded, and reasoning is strictly monotonic. Logic programming does not readily provide means for handling defaults and exceptions as frames do.

Production Rules

The most widely used schemes for knowledge representation are rule-based systems (Harmon and King, 1985). Among the architectures of rule-based systems, the production system architecture has been most popular (Brownston *et al.*, 1985). It is a simplified version of the blackboard architecture discussed below. A production system consists of a knowledge base and an inference engine. The knowledge base consists of data (objects, facts, goals) and rules. Data are stored in the working memory, which serves as a global database of symbols representing facts and assertions. The set of rules constitute the program's core, which is stored in the production memory (Winston, 1984). Each rule has a condition part, which consists of one or more antecedent clauses, and an action part, the consequent, which may create or modify working memory elements. The inference engine of the system determines which rule instantiations are relevant to a given working memory configuration and assembles them in a conflict set. The inference engine then uses a conflict resolution strategy to fire one or more rules from the conflict set. Some of the strategies followed for selection of a rule are recency of the working memory element involved in the condition part of the rule, and specificity of the rule measured by the number of condition elements (Brownston *et al.*, 1985).

There are two ways that rules can be used in a rule-based system: forward chaining (OPS5 like) and backward chaining (PROLOG like). In forward chaining, rules are matched against facts to establish new facts or hypotheses. In backward chaining, the system starts with what it wants to prove and tries to establish the facts it needs to prove it. This matching of rule conditions to the facts produces what are called "inference chains." The inference chains indicate how the system used the rule to make the inference, and they can be represented in an AND/OR tree. The methods of selecting and firing rules in forward-chaining rules is different from those of backward chaining rules.

Rules are appropriate for image interpretation because a major part of domain-specific knowledge results from empirical associations (heuristics) developed through years of experience in a particular area, which may be expressed as heuristic rules. However, representing knowledge as an unordered and unstructured set of rules has certain disadvantages. For example, one cannot easily express the structure of the domain in terms of taxonomic, or part-whole, relations that hold between objects and between classes of objects. Frames provide for such a mechanism.

Frame Systems

Frames (scripts of schemata) are structural models for representing stereotyped objects or situations (Minsky, 1975). A class frame is a collection of all the relevant information that describes a class of objects. An object or instance frame is a collection of all the relevant information that describes an individual or instance of a class frame. A frame has slots that contain properties and relationship information about classes and objects as well as procedural attachments. The slots specify what we expect to know about an object class and how we expect to acquire it. The number and type of slots are fixed when the frame is defined. Procedural attachments or active values are rule sets or procedures associated with the slot of a frame which must be invoked before a value can be assigned to or read from that slot. Thus, they behave like "demons" monitoring changes and uses of the values. These attachments may contain links to other frames, slot values, predicates, procedures, and hypotheses or conjectures about other objects or classes. Frames can control the invocation of knowledge sources (large rule sets) within the frame context of a blackboard system. Once a frame is hypothesized, it is invoked or instantiated as a frame instance, that is, the slots are filled to determine if the frame matches the object being considered.

The hierarchies of a vision system as described by Kanade (1977) can be expressed in a frame system in an explicit manner. These hierarchies are a processing unit hierarchy, a detail hierarchy, and a composition hierarchy (Figure 5). Within the image domain the processing unit hierarchy refers to the size of the areas processed in an image such as pixels or regions. The detail hierarchy, which is in both the scene and image domains, refers to the precision of description. The description may range from crude to very detailed. In the scene domain this might be the precision of shape descriptors. In the image domain this may be the resolution of the data. The composition hierarchy includes part-of relations, such as object to subobject designations. Complex objects are represented as compositions of simpler objects, thus creating a composition hierarchy. The recognition of a complex object involves the recognition of its component parts such that the constraints of the frame are satisfied.

Frames can contain both declarative (property values) and procedural knowledge (procedural attachments) which is important in a combination top-down and botton-up control strategy. The procedural information is useful for guiding the search process for instances of the object class. A procedural approach is most useful when the structure of the image is well known so that the operators and their sequencing can be determined. This is the same class of scenes in which a top-down approach is most useful. When the scenes have great variablility in structure, then the analysis process will be more data driven. In this situation, large numbers of variables will have to be computed and verified, which complicates the control in a procedurally oriented process. A declarative approach will have advantages in this situation. In a top-down search a frame's hypothesis provide parameters to the subframes. In a bottom-up search the hypothesis expectations can restrict the subframes which can invoke the frame as a supergoal. Verification is done by matching of the knowledge base to the data. One uses image features computed from the input data to find candidate objects, and then one attempts to match the expectations of each candidate object to the image features. That is, one uses image features to reference into a frame which limits the expectation of the next step in the analysis. This allows one to selectively activate knowledge.

Because of their structure, frame systems are useful for image interpretation because expectations about the form, relations, and recognition procedures of the objects play an important role in image interpretation.

Blackboard Architectures

In the organization of a "blackboard system architecture" (Figure 5), each node in the object hierarchy has a package of "how-to" knowledge, specific to that node, which is often represented in the form of a cluster of rules called knowledge sources (Engelmore and Morgan, 1988). Usually, there is no need for a uniform mechanism which operates on some description of the knowledge within each knowledge source, but, rather, each knowledge source has different kinds of knowledge and different mechanisms for carrying out its reasoning (Chandrasekaran, 1983). The generated solution elements by the independent knowledge sources are recorded on the blackboard or short term memory (STM). The "blackboard," like the working memory of a production system, is a global database for recording solution elements generated during problem solving (Figure 5). At any given time the several knowledge sources that contribute to the solution of the problem communicate only by writing on the blackboard. Blackboards have mechanisms associated with them for invoking pattern directed procedures (active values or demons) and for

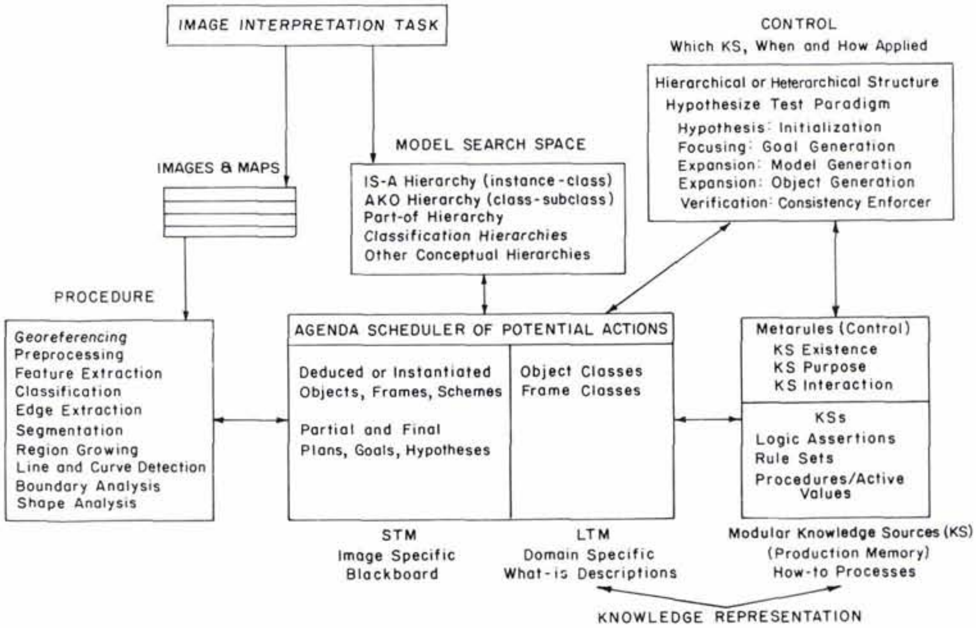

FIG. 5. A blackboard architecture model for image interpretation.

synchronizing their activities (Hanson and Riseman, 1978). Often potential actions awaiting execution are recorded in the agenda of the blackboard. A scheduler can be used for selecting the knowledge source to be executed at a given time by determining which pending action from those stored in the agenda should be executed next. A consistency enforcer maintains a consistent representation of the emerging solutions (Hayes-Roth *et al.*, 1983). This may take the form of likelihood revisions or a belief maintenance system that combines the information obtained from the different knowledge sources. Several belief maintenance systems have been proposed; some are heuristic and some have a theoretical basis. Fuzzy logic (Zadeh, 1978) and the Dempster-Shafer theory (Shafer, 1976) are examples of systems with a theoretical base.

The distribution of knowledge among various specialists (knowledge sources) modularizes the knowledge base in a natural way. The criteria for distribution, that is, deciding which piece of knowledge should go in what knowledge source, is problem dependent. Separation of the various types of knowledge in separate knowledge sources is desirable because it provides modularity; however, it does complicate the problem of control and communication.

INFERENCE AND CONTROL STRATEGIES FOR IMAGE SEGMENTATION

A knowledge-based image interpretation system can be conceptualized as being composed of preprocessing, segmentation and classification procedures, a knowledge base of models and how-to knowledge, a database of facts and hypotheses pertaining to the current image, and an inference and control system (Figure 5). The segmentation and classification procedures are used to extract image features and to label and describe mostly edges and regions. The role of the inference system is to sequence or select the models in the knowledge base, match the extracted image features against the models, resolve conflicts, and track the inferences. The function of the control portion of the inference engine is to decide how to start the reasoning process with all its facts and rules, and how to resolve conflicts and decide which rules to fire among conflicting rules that are ready to fire. Three control strategies are often employed: top-down, bottom-up, and a combination of the first two strategies. This terminology has its origin and was adapted from grammatical pattern recognition and formal language theory in analogy to grammatical parsing that takes place in a top-down or bottom-up manner. In a top-down approach, one starts with an assumption or hypothesis about the objects on the image and their appearance. Each hypothesis is then decomposed into subhypotheses and so on. The lowest level hypotheses must then be verified by extracting image features. Matches are made at lower levels of the hierarchy to determine if the hypotheses are valid. If the hypotheses are not valid, another hypothesis is chosen at the higher level of the hierarchy.

Top-down processing is goal directed processing. High-level goals generate subgoals until one can solve the goals and then back the results up the hierarchy. A number of image interpretation systems that employ a top-down stategy have been developed to analyze chest radiographs (Harlow, 1973; Harlow *et al.*, 1976). A top-down approach works for this class of scenes because they are highly structured, that is, there is available certain *a priori* information about the expected image objects and their relations. If *a priori* or auxiliary information is available about the scenes of concern, then one can presumably generate a set of meaningful goals or hypotheses to pursue, and a top-down approach is appropriate. Top-down analysis is similar in spirit to backward-chaining in an inferencing system which proves goals by seeking to prove their subgoals. In general, a pure top-down approach is not suitable for images with a large amount of variablility in their structure such as aerial images.

Most image interpretation systems use a bottom-up strategy, a process which begins at the bottom of the hierarchy with the identification and examination of fine detail, e.g., cars and houses, and attempts to identify more general entities such as residential areas by recognizing appropriate groupings of the smaller objects. One usually starts by segmenting the image to find homogeneous regions or lines that correspond to objects at the lowest level. Local properties are usually used to produce the low-to medium-level segmentation, and no resource to semantics is made. Simple properties of scene geometry and gray level contrast between object and background typify the process (Ballard and Brown, 1982; Marr, 1982). The low- and medium-level segmentation process is in general difficult to perform correctly. Even if a good segmentation is achieved, the segmented image often contains less information than the original image. The lost information might be important in obtaining the best high-level interpretation of the scene. The analysis is driven by the data at the lowest level. Bottom-up analysis is similar in spirit to forward chaining in an inferencing system which derives consequences from established results.

A bottom-up approach has the following limitations:

- The segmentation of the scene is based entirely upon image features obtained from segmentation algorithms which have relevance only to the information at the lowest levels of the hierarchy in the scene domain. These operators use little or no information from the scene domain, which makes arbitrary interpretations and errors likely. Because little context is utilized, an extensive search process is required to locate regions which might correspond to the low-level objects in the scene domain and for matching all the low-level objects to synthesize high-level objects.
- The objects in the scene domain which can be explicitly verified in the image domain correspond to low-level objects. This means complex models in the scene cannot be directly verified, and independent evidence cannot be gathered at intermediate levels of the hierarchy for a belief maintenance system.
- Because explicit models for high-level objects and their relationships cannot be verified with image features, the objects at high levels in the scene domain hierarchy must be verified by defining relationships on lower level objects. After one has obtained interpretations of low-level objects from image ques, one can then infer the interpretations of the objects at higher levels in the hierarchy.

In general, it is important not to constrain a vision system to a pure top-down or bottom-up approach. One needs a vision system which incorporates both goal- and data-driven analysis. Havens and Mackworth (1983) developed a hierarchical scene analysis system to overcome the limitations of bottom-up and top-down approaches. At each node of the hierarchy a scheme or frame was proposed to model each abstract object. This system provided for procedural knowledge at each node to guide the search for instances of objects. Nagao and Matsuyama (1980) suggested a trial and error recursion between bottom-up segmentation and the top-down recognition to reach the best results.

While, for simple images, hierarchical top-down or bottom-up approaches may do, more complex images—like high-resolution remotely sensed data—may require heterarchical processing rather than strictly hierarchical processing (Narasimhan, 1974). Heterarchical processing involves the possibility of cycling back through a hierarchy to aid further discrimination of objects (Ballard and Brown, 1982). Heterarchic control structures are suitable for complex image interpretation tasks because such tasks can be decomposed into smaller subtasks, each requiring some expertise, and are not necessarily performed globally in a fixed order. A community of cooperating and competing specialists (knowledge sources) and a blackboard represents such a concept (Figure 5). In such a system, the expert who can help most toward the final task solution at any given time is selected. It is selected because according to some criterion its subtask is the best thing to do at that time. In "opportunistic problem solving," cooperation among the knowledge specialists is achieved by assuming that whatever information is needed is supplied by someone else. As new pieces of evidence are found and new hypotheses are generated, appropriate knowledge sources analyze them and create new hypotheses. Each specialist must rely on the other specialists to supply the information each needs. The criteria for selection are wide and varied and several ideas have been tried (Hanson and Riseman, 1978; Nagao and Matsuyama, 1980; Nicolin and Gabler, 1987). Control may be provided by a single executive (scheduler), by metarules (e.g., a universal set of rules that controls the invocation of other rules), or by an *a priori* system of ranking (Ballard and Brown, 1982). Practically, only parallel computer architectures could run cooperative specialists, yet this process has been simulated in a serial machine (Hanson and Riseman, 1978).

KNOWLEDGE SOURCES AND TYPES FOR IMAGE INTERPRETATION

The traditional approach to processing remotely sensed data for thematic map creation involves acquiring remotely sensed data; georeferencing that data; integrating newly acquired data within a geographic information system (GIS); classifying, extracting attributes, and labeling image objects utilizing the image analysis or computer vision system and the GIS; updating the GIS; and producing output thematic maps. The knowledge required to perform these functions is expressed in either declarative or procedural form. Known facts and relationships are declarative knowledge. Currently, the required declarative knowledge is provided by expert photointerpreters utilizing domain specific knowledge and their perceptual skills. Procedural knowledge is most often expressed as programs and describes how to locate, recognize, and classify features. There is a substantial amount of procedural knowledge that has been accumulated. Image analysis and GIS systems are comprehensive, encompassing thousands of lines of FORTRAN or C code. The GIS systems are equally large and complex. There is substantial processing required in the image analysis and pattern recognition algorithms; the GIS storage, retrieval, and analysis functions; and the conversion functions between different projections and types of data. In image analysis, for example, there are issues relating to the interaction between the segmentation algorithms and the inferencing system, the selection and adequacy of the segmentation algorithms, the method of forming an optimal global interpretation of the scene, and the integration of system modules and data structures so that the system can be improved as additional knowledge about the scene becomes known. The correct sequencing and selection of these functions requires substantial knowledge for processing, analysis, synthesis, and presentation of the data in a manner which will yield best interpretation and resultant maps (Schowengerdt and Wang, 1989).

The use of domain knowledge is critical to the development of robust systems for automated mapping of Earth features. Domain knowledge is designated in the following as general knowledge, discipline knowledge, regional knowledge, and object knowledge. The manner in which knowledge can be structured and utilized in the analysis of images is an important consideration. In this section some means in which knowledge can be used to guide the interpretation system are indicated.

General knowledge includes information about the imaging system and processes by which the images were acquired (ephemeris data)-including image conditions, such as latitude, longitude, date, sensor type, calibration parameters, and atmospheric conditions—and general knowledge about the types of objects expected in the given image. By combining location with global environmental, geological, and political maps, *a priori* information such as large ecological communities, physiographic regions and sections, and the country and the county could easily be determined. Knowing the country or county is important. For example, grain silos in the United States are round, while those in Canada are square. Further, the distance to the nearest urban area would give an indication of whether the scene was part of a natural or manmade landscape. Knowing the physiographic section, one could infer the possible landforms found in the region. The date can also provide useful information, because the objects one expects to find in a scene are often dependent on the time of year. For example, it would be futile to use the attribute "has leaves" to identify deciduous trees if the image was obtained during the winter. Further, winter scenes could contain snow and ice, whereas summer scenes would not.

Discipline specific knowledge includes information about the spectral, temporal, and structural properties of objects, such as their reflectance characteristics, relative size, height, shape, texture, and their decomposition and embedding in other objects; and information about contextual and semantic constraints among objects, such as site specification (relative location) and associations with other objects or phenomena. Having knowledge concerning the basic physical or theoretical principles in a particular discipline increases the ability to accurately identify and map specific features. For example, if forests are to be mapped, then a general photointerpreter could do the job. If detailed communities within the forest are to be mapped, a forestry background would be required. In mapping forest communities, certain desired classes would need to be merged or separated depending on the season of image acquistion. Forest communities also are generally located in specific spatial arrangements.

Regional knowledge about the geographic area represented in the image includes existing maps and reports of various kinds, literature sources, case studies, and communication with experts related to specific tasks at hand. Regional knowledge is needed to capture region-specific details that take into account physical, cultural, or temporal variations that deviate from the expected conditions. Although similar environments contain many similar objects, many objects are region dependent and, therefore, regional knowledge adds a better definition to the discipline knowledge. Furthermore, site-specific knowledge or human expertise is required to refine the reasoning relative to local exceptions or unique cases learned from experience. For example, a marsh in Louisiana or California could be expected to have oil wells, whereas this would not be expected in a New England marsh. An oil processing facility on the coast is much more likely to be located adjacent to a canal than in the middle of a marsh. If one was searching for such a facility, then the search should first focus on areas near canals. Similarly, an object located greater than 100m from a canal would probably not be an oil facility.

A particular object can have many spatial attributes. However, some of these attributes are more powerful for identification purposes than others. For example, shape is not a particularly useful attribute for identifying the World Trade Center, whereas height is. It is therefore extremely useful to know the "most distinguishing characteristic" of on object. This characteristic is not absolute, however, but depends on context. For example, height would not be the most distinguishing characteristic for a tall building in a city full of skyscrapers. The most distinguishing characteristic must be derived given the context of the specific area. The function of an object can be used to locate and identify the object is examined. Many objects have a wide variety of possible forms or shapes in which they may appear in imagery, which makes it difficult to characterize them by shape or other properties. Road networks have widely varing geometries and sizes for example. The purpose of the road network is to provide access to cities, residential areas, etc. A similar statement is true for canals that provide access to petroleum exploration platforms. Buildings and plants have widely varying shapes. If the plant produces electricity or petroleum products, there will be basic differences in the two facilities because of their function. The function of an object can be studied to gain additional insights into determining distinguishing charachteristics of objects and the manner in which these characteristics can be used by the interpretation system. Another factor related to objects is that certain features add more to our understanding of a scene. Locating these objects initially, therefore, aids in subsequent identification. For example, in a scene with motels, amusement parks, jetties, and piers, the object that would add most to our contextual understanding might be the beach.

KNOWLEDGE-BASED VISION SYSTEMS IN REMOTE SENSING

Expert interpretation systems have been developed to recognize objects primarily from high spatial resolution black-and-white images. Some aspects of these systems are discussed. Emphasis is on computer vision and knowledge-based techniques employed for the interpretation of aerial images. No attempt is made to provide a complete survey. Most of the systems presented have been described in multiple papers and reports, only a few of which are cited in this paper. The interested reader may wish to contact the system designers for current system design and status.

CERBERUS was developed to structure rules pertaining to the spectral responses in MSS channels 5 and 7 and ancillary data such as elevation, slope, and prior land cover (Engle, 1985). Fifty-five rules were used to distinguish among five of the seven level I land-cover types. CHESHIRE, an enhanced version of CERBERUS, opperates in the Xerox 1108 LISP Dandelion workstation. This methodology was orignially developed by Erickson and Likens (1984). They defined Landsat MSS taxonomies, through semantic nets, defining important terms and relationships within auxiliary data. Rule sets were constructed for urban, agriculture, range, forest, water, bare classes, and a final summary rule set. Contingency tables were used to represent relationships between spectral classes and values in the ancillary data.

Wharton (1987) demonstrated a land-cover classification rule-based system that utilized the relationships between the spectral values of adjacent pixels as described by expert photointerpreters instead of training samples. Band-to-band, category-to-category, and category-to-background relationships are used to quantify the color-contrast features of the knowledge base. A hierarchical data structure was used to compute contrast values between neighboring pixels at various levels of detail. The methodology was demonstrated for land-cover classification from high resolution TM simulation data.

Civco (1989) designed an expert system for Level I and Level II land-use mapping by employing expert image analysis rules and heuristics to classify Landsat TM data. Knowledge from spectral, spatial and temporal domains was addressed. Physical principles, expert intuition, and inference induction were employed for knowledge acquisition. The rule-based system was developed in the expert system tool EXSYS. The system contained 94 Level I and 49 Level II classification rules. The expert system read its input data through some record structures produced from the image analysis systems. A comparison has indicated that the results were superior to those achieved through supervised per-pixel classification.

VISIONS is a computer system for interpreting natural scenes (Hanson and Riseman, 1978). VISIONS includes two distinct parallel iterative segmentation algorithms: the first aggregates edges into boundaries while the second utilizes global histograms and a local spatial analysis procedure to form regions. VISIONS computes and symbolically represents regions, boundary segments, and two-dimensional shape attributes. Multilevel structures are used for representing the model being built (short term memory) and the stored world model (long term memory). Schemas (frames) classes and instances represent objects in the scene. The nodes of the abstract hierarchy include objects, volumes, surfaces, regions, segments, and vertices. Control strategies decide which partial model in the model search space needs to expand, which level of representation to select, and which hypotheses at that level need to expand. The specific processes are focusing for goal generation, expansion for models and object generation, filtering for hypothesis elimination, and verification (Figure 5).

Nagao and Matsuyamam (1980) described an expert vision system for classification of multispectral data from suburban scenes. Characteristic regions are a focal point of the analysis and are defined by spectral and spatial features such as color, size, shape, and texture. Examples would be large homogeneous, elongated, shadow, water, and vegetated regions. Spatial characteristics of these regions have provided more consistent results than spectral properties, especially in cases of changes in imaging conditions. A production-like system is used to incorporate knowledge in the system. It employs production rules and image features for the image extracted at a low-level. Heuristics are expressed in locational constraints and spatial arrangement rules. The choice of a production system architecture was made because of its control mechanism. The authors felt that the adaptive activation of rules available in the production system architecture was best suited for a generic system that can not assume or anticipate the instantiated objects of a given scene. Facts and events were stored in a blackboard which was accessed by independent knowledge sources which performed object detection. An aerial scene analysis system called SIGMA is described by Matsuyama (1987) and by Matsuyama and Hwang (1987). The system utilizes frames and is based on a blackboard model for uniform communication among independent specialist modules. SIGMA exhibits mechanisms for the focus of attention, conflict resolution, and the correction of early segmentation errors.

Levine and Shaheen (1981) describe a system applied to natural scenes. The modules of the system are low-level processes, measure analyzer, hypothesis initializer, hypothesis verifier, focus of attention, and scheduler. These processes communicate through long-term memory and short-term memory similar to the blackboard architecture. The low-level processes give simple image segmentation. The measure analyzer computes measures over regions and other structures. The hypothesis initializer uses region descriptions and model information in long-term memory to generate interpretations. The hypothesis verifier uses measures from regions, relations between regions, and hypotheses about regions to verify and update interpretations. The focus of attention module recognizes situations of interest and generates actions. The scheduler controls execution of modules. The data are arranged in a relational database, and the system is implemented as a rule-based system. Nazif and Levine (1984) used a rule-based system for low-level segmentation into uniform regions and connected lines. A focusing mechanism was employed to concentrate on "interesting" parts of the image.

Binford (1982) reviewed computer vision systems and described the development of the ACRONYM system. Points are made in this article that most systems operate on two-dimensional data, use fairly simple world models, have limited segmentation procedures, and use only weak descriptors of shape and texture (Brooks, 1983). The ACRONYM system uses generalized cylinders for description of 3-D shapes. The models provided a representation which is independent of viewing angle. The user describes the classes to be interpreted in the image and their spatial relationships to other classes and subsets of those classes. The system first deduces the volumetric models from the descriptions of the users and then labels their objects in the image for which it has a consistent interpretation. The recognition strategy is bottom-up.

McKeown *et al.* (1985) described a rule-based system for interpretation of airport scenes. The rule-based system interprets the scene by building interpretations based upon an initial segmentation which is produced by a region growing program. Region properties are extracted to determine an association between regions and airport features. An initial confidence is calculated to detrmine how well a region fits the feature description. Rules are organized into classes. Initialization rules determine the goal states, map database, class expectations, and low-level segmentation. Region-to-interpretation rules create an initial hypothesis for each region. Local evaluation rules are used to enlarge regions. Consistency rules apply spatial and context constraints to modify the confidence of initial fragment hypotheses. Functional area rules recongize when fragment interpretations can be grouped into functional areas such as runways, taxiways, and tarmac. Goal generation rules recognize situations inconsistent with airport structure in order to prune weak fragments from further consideration. Model generation rules assemble functional areas into a model for the airport scene. McKeown (1987) employed a flexible interface for a GIS that allows querying by pointing to an image or map display.

Nicolin and Gabler (1987) designed an expert system for interpretation of suburban scenes from low-altitude aerial photographs. A semantic network represented declarative knowledge in their system. The semantic network is structured by two hierarchies of relations. The first hierarchy is the generalization and specialization relation. Inheritance occurs through specialization. The other heirarchy is by the composition and decomposition relation. This hierarchy gives the structure of complex objects from less complex objects. The system provides for long-term memory and short-term memory. Long-term memory contains the generic knowledge of the semantic net. Short-term memory contains intermediate and final results of processing steps. Processing is carried out by a series of modules. Low-level modules determine bright, dark, and border areas. Medium-level modules incorporate segmentation algorithms. High-level modules perform object identification. The control mechanism is bidirectional: the data-driven (bottom-up) algorithms are used when there is no active hypothesis to direct the location of elementary components; the model-driven (top-down) mechanism takes over as soon as a sufficient number of image fragments have been identified to form a hypothesis.

Goldberg *et al.* (1983) have described a production rule-based system for integrating multitemporal Landsat images for classification of forested areas. After a newly acquired image was classified, rules based upon the present classification, the previous classification, and various measures of confidence are invoked by the data and a new decision is computed for each pixel. The production rules were supplied by experts. Improvements in the classification of subtle forest species were observed. Automatic estimation of forest depletion by logging was modeled by Goldberg *et al.* (1985). The system is decomposed into a number of specialist experts organized in a hierarchical fashion around a series of blackboards which are used for communications between the different levels. The lowest level expert provides the interface to the image processing algorithms. Other experts included an expert for cloud-and-shadow determination, an expert for change detection, and an expert in maps and geocoded databases. Goodenough *et al.* (1987) have designed an expert system in PROLOG that contains objects and a metarule interpreter, a blackboard for intermediate results, a scheduler, an explanation facility, and a contention arbitrator. Control is both forward and backward. The low-level image analysis takes place in FORTRAN. One of their applications enables analysts to choose suitable features on classified TM raster images for matching against the stored vector based GIS database. Selected image segments are transformed to bring them into congruence with the map. Another expert system helps the non-computer specialist to use all the tools of the system to perform a given analysis task such as selecting suitable training sites for spectrol classifications.

Argialas and Narasimhan (1988a) have designed a rule-based expert system, the Terrain Analysis Expert (TAX). Knowledge pertaining to the landform-pattern element approach was represented in physiographic section models and landform models. The physiographic section models represented the relations among the sections and the landforms that can occur in them. The landform models contained information about all the pattern element values that were likely to be found in a landform, and the likelihood of their occurrence. The system queried the user for the certainties of the pattern element values of an unknown site, fired appropriate rules, and reported the type of landform that best matched the pattern elements of the unknown site. TAX was implemented in the production system architecture of the OPS5 language (Argialas and Narasimhan, 1988b. The domain-specific knowledge about terrain analysis was separated into two components; one component consisting of specific knowledge about landforms, stored as facts in the working memory, and the other component consisting of the general methodology for reasoning, stored as rules in the production memory. To handle the uncertainties introduced during problem solving in both the identification of the individual pattern elements and the synthesis of the pattern elements in inferring the landform of the site, it was judged appropriate to associate certainty values with each pattern element value observed on an aerial image and employ them in decision making. Moreover, probability values were associated with each fact in the models of landforms to express its strength in the identification of a particular landform type. Fuzzy sets and the Dempster-Shafer theory of evidence have been applied for representing pattern element values and combination of evidences (Narasimhan and Argialas, 1989).

A frame-based model has been designed for knowledge representation and problem solving in terrain analysis (Argialas, 1989a). Frames were developed to represent relations between physiographic sections and landforms, landforms and their pattern elements, and pattern elements and their associated likelihood of occurrence in each landform type. Frames have been used to demonstrate inheritance of attributes from generic representations of terrain units to their specific instantiations. Frames have also been used to represent procedural knowledge by embedding such knowledge in the form of attached predicates. The methodology demonstrated the representation and reasoning capabilities of frames, backward and forward chaining rules, and inexact reasoning for the interpretation of landforms from aerial images. The Terrain Analysis Expert-2 (TAX-2) was implemented in the frame- and rule-based expert system tool called Intelligence Compiler (IntelligenceWare, 1986). Class frames, object frames, attributes, and values are a natural way to represent structure in terrain analysis, as is the specification of inheritance hierarchies among terrain objects.

Mintzer (1989) designed an expert system for landform identification. An expert system shell—Knowledge Engineering System (KES)—was used to encode domain-specific knowledge relating photo-identifiable features to specific landforms. The resulting interactive software system identifies individual landforms observed in stereo aerial photography based on a set of key pattern elements elicited from the terrain analyst. This system contains a large knowledge base, and it is almost in the stage of an operational system.

Interpretation systems have progressed so that the inferencing and control mechanisms are quite complex. The segmentation algorithms used with image interpretation systems for the most part have changed little. There has been improvement in obtaining interpretation of the outputs of these segmentation algorithms with better inferencing systems. It is clear that, to achieve the analysis of complex aerial scenes, substantially more complex systems are needed.

AN OVERVIEW AND OUTLOOK

Knowledge-based image interpretation can upgrade the state of image analysis capabilities from brute force mathematical and statistical approaches to analysis techniques based on interpretation logic and heuristics. Statistical and analytical algorithms will be applicable in expert interpretation systems, but as low- and medium-level labeling techniques selected and controlled by conceptual reasoning. In creating an image interpretation system, a number of problems must be addressed: the selection and adequacy of the segmentation and classification algorithms, the interaction between the segmentation/classification algorithms and the inferencing system, the method of forming an optimal global interpretation of a scene, and the integration of system modules and data structures so that the system can be improved as additional knowledge about the scene becomes known (Figure 4). There is no general theory for selecting the measurements, features, description, representation, segmentation, recognition, and classification techniques needed for the implementation of generic interpretation systems. Although some aspects of these problems have elegant theoretical formulation, the state-of-the-art is strictly problem dependent. Heuristic features are largely responsible for almost all the practical pattern recognition systems to date. Feature and attribute selection rely on the past experience, engineering intuition, and domain specific knowledge of the designer. One can only hope to select some of the possible discriminatory features or attributes. Attribute selection processes may be validated but are not easily optimized. Segmentation and classification methods are selected and evaluated in light of their performance in a given application based on experimentation and judgment. There is a lack of universal and context independent segmentation and classification techniques. The choice of one segmentation or classification technique over another is dictated mostly by the peculiar characteristics of the problem being considered. At this point it seems that task dependent approaches are necessary for high level image interpretation. In most cases, the designer constructs the hierarchical/relational organization, semantic net, or production grammar based on domain knowledge and his experience (heuristic knowledge). For successful interpretations, very detailed specific knowledge of the scene being analyzed is required. Automatic inference or induction techniques are needed to assist knowledge elicitation. It is likely that in the foreseeable future the human will be a part of the analysis, and one must consider augmenting his functions with reliable image analysis components as they become available.

The analysis and interpretation of remotely sensed images at a high level of detail is a complex task. Although some researchers have produced promising results, more research is required before interpretation systems can be usefully and cost-effectively applied to problem solving in remote sensing. Substantial research is required to define how image interpreters perform their job and to formalize this process before it can be automated. If we cannot formalize how analysts go about their tasks, we cannot automate their procedures. It seems that funding is highly justifiable in pursuing these research issues. From an educational point of view, computer vision and expert system courses need to be integrated in to remote sensing, mapping, and computer and information management curricular (Argialas, 1989b). Students trained through such curricula will be at the core of the professional community that will implement and use such hybrid geo-information systems.

ACKNOWLEDGMENTS

The authors are thankful to Nicholas Alvertos, William J. Campbell, Robert F. Cromp, Eric Dorfman, Andrew Frank, Jeffery L. Kretsch, Sankar Krishnamurthy, Scott Leibowitz, John G. Lyon, Ravi Narasimhan, and Robert A. Schowengerdt for providing a critical review and valuable editiorial and organizational suggests. Their comments and suggestions improved the quality of the paper.

REFERENCES

Ahuja, N., and B. Schachter, 1983. *Pattern Models*, John Willey and Sons, New York, 309 p.

Alvertos, N., D. Brazakovic, and R. Gonzalez, 1989. Camera Geometries for Image Matching in 3-D Machine Vision, *IEEE transactions on Pattern Analysis and Machine Intelligence*, PAMI-11, No. 9, pp. 897–915.

Argialas, D. 1989a. A Frame-Based Approach to Modeling Terrain Analysis Knowledge. *Proceedings Annual Convention of American Society for Photogrammetry and Remote Sensing*, Vol. 3, pp. 311–319, 2–7 April 1989, Baltimore, Maryland.

———, 1989b. Teaching Expert Systems Techniques at Louisana State University. Chapter 3 in *American Society of Civil Engineers Monograph: Expert Systems for Civil Engineers-Education*, (S. Mohan and M. Maher, editors), American Society of Civil Engineers (ASCE), pp. 23–29.

Argialas, D., J. Lyon, and O. Mintzer, 1988 Quantitative Description and Classification of Eight Drainage Pattern Types. *Photogrammetric Engineering & Remote Sensing*, Vol. 54, No. 4, pp. 505–509.

Argialas, D., and R. Narasimhan, 1988a. TAX: A Prototype Expert System for Terrain Analysis. *Journal of Aerospace Engineering*, Vol. I, No. 3, pp. 151–170.

———, 1988b. A Production System Model for Terrain Analysis Knowledge Representation. *Microcomputers in Civil Engineering*, Vol. 3, No. 1., pp. 55–73.

Ballard, D., and C. Brown, 1982. *Computer Vision*, Prentice-Hall.

Bezdek, J., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 256.

Binford, T., 1982. Survey of Model-Based Image Analysis Systems, *International Journal of Robotics Research*, Vol. 1, No. 1, pp. 18–64.

Brayer, J., P. Swain, and K. Fu, 1977. Modeling of Earth Resources Satellite Data, *Syntactic Pattern Recognition Applications, K. Fu editor) Springer-Verlag, Berlin*, pp. 215–242.

Brooks, M., 1978. Rationalizing Edge Detectors: Some Comparisons, *Computer Graphics and Image Processing*, Vol. 8, pp. 277–285.

———, 1983. Model-based Three-Dimensional Interpretations of Two-Dimensional Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAM 1-5, pp. 140–150.

Brownston, L., R. Farrell, N. Martin, and E. Kant, 1985. *Programming Expert Systems in OPS5*, Addison-Wesley.

Campbell, W. J., and L. Roelofs 1984. Artificial Intelligence Applications Concepts for the Remote Sensing and Earth Science Community: *Proceedings, IX Ninth Pecora Symposium on Spatial Information Technologies for Remote Sensing Today and Tomorrow*, Sioux Falls, South Dakota, pp. 232–242.

Chandrasekaran, B., 1983. Towards a Taxonomy of Problem-Solving Types, *AI Magazine*, Vol. 4, pp. 9–17.

Chen, P., and T. Pavlidis, 1979. Segmentation by Texture Using a Coocccurrence Matrix and Split-and-Merge Algorithm, *Computer Graphics and Image Processing*, Vol. 10, pp. 172–183.

Civco, D., 1989. Knowledge-Based Land Use and Land Cover Mapping, *Proceedings, Annual Convention of American Society for Photogrammetry and Remote Sensing*, Vol. 3, pp. 276–291, 2–7 April 1989, Baltimore, Maryland.

Clowes, M., 1971. On Seeing Things, *Artificial Intelligence*, Vol. 2, No. 1, pp. 79–112.

Conners, R., M. Trivedi and C. Harlow, 1984. Segmentation of a High Resolution Urban Scene Using Texture Operators, Computer Vision, *Graphics and Image Processing*, Vol. 25, pp. 273–310.

Duda, R., and P. Hart, 1973. *Pattern Classification and Scene Analysis*, Wiley, New York, N.Y. 482 p.

Engelmore, R., and T. Morgan, 1988. *Blackboard Systems*. Addison-Wesley, New York, N.Y., 602 p.

Engle, S., 1985. The Development of an MSS Satellite Imagery Classification Expert System, DIGEST, *Proc. 1985 Intl. Geoscience and Remote Sensing Symp (IGARSS '85)*, 7–9 Oct. 1985, Amherst, Mass., Vol. 2, pp. 693–696.

Erickson, W., and W. Likens, 1984. An Application of Expert Systems Technology to Remotely Sensed Image Analysis, *Proc. IX Pecora Symposium on Spatial Information Technologies for Remote Sensing Today and Tomorrow*, pp. 258–277.

Estes, J., E. Hajic, and L. Tinney, 1983. Manual and Digital Analysis in the Visible and Infrared Regions, *Manual of Remote Sensing*, Second Edition, American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, Vol. 1, pp. 987–1123.

Estes, J., C. Sailor, and L. Tinney. 1986. Applications of Artificial Intelligence Techniques to Remote Sensing, *Professional Geographer*, Vol. 38, pp. 133–141.

Fabbri, A. G., K. Fung, and S.M. Yatabe, 1986. The Usage of Artifical Intelligence in Remote Sensing: A Review of Applications and Current Research, *Proc. NATO Adv. Study Inst. on Statistical Treatments for Estimation of Mineral and Energy Resources* (C.F. Chung, A. G. Fabbri, and R. Sinding-Larsen, editors), Reidel Publishing Dordrecth.

Findler, N., 1979. *Associative Networks*, Academic Press, Orlando, Florida.

Friedl, M., J. Estes, and J. Star, 1988. Advanced Information-Extraction Tools in Remote Sensing for Earth Science Applications: AI and GIS. *AI Applications* Vol. 2, pp. 17–31.

Fisher, P., and W. Mackaness, 1987. Are Cartographic Expert Systems Possible? *Proceedings, Eighth International Symposium on Automated Cartography*, Baltimore, Maryland, pp. 530–534.

Fu, K., 1974. *Syntactic Methods in Pattern Recognition*, Academic Press, New York, 296 p.

———, 1976. Tree Languages and Syntactic Pattern Recognition, *Pattern Recognition and Artificial Intelligence* (C. Chen, editor), Academic Press, New York, pp. 257–291.

———(editor), 1977. *Syntactic Pattern Recognition, Applications*, Springer-Verlag, Berlin, 270 p.

———, 1980a. Syntactic (Linguistic) Pattern Recognition, *Digital Pattern Recognition*, (K. Fu, editor), Springer-Verlag, Berlin, Chapter 4, pp. 95–134.

———(editor), 1980b. *Digital Pattern Recognition*, Spring-Verlag, Berlin, 296 p.

———, 1981. Recent Progress in Syntactic Pattern Recognition, *Progress in Pattern Recognition* (L. Kanal and A. Rosenfeld, editors), North-Holland, Amsterdam, Vol. 1, pp. 1–13.

———, 1982a. Attributed Grammars for Pattern Recognition-A General (Syntactic-Semantic) Approach, *IEEE Press Proceedings, Conference on Pattern Recognition and Image Processing*, IEEE Computer Society, Las Vegas, Nevada.

———, 1982b. Hybrid Approaches to Pattern Recognition, *Pattern Recognition Theory and Applications* (J. Kittler, K. Fu, and L. Pau, editors), D. Reidel Publishing Company, Dordrecht, Holland, pp. 139–155.

———(editor), 1982c. *Applications of Pattern Recognition*, CRC Press, Boca Raton, Florida, 2177 p.

Fu, K., and P. Swain, 1971. On Syntactic Pattern Recognition, *Software Engineering*, Vol. 2, pp. 155–182.

Fu, K., and A. Rosenfeld, 1976. Pattern Recognition and Image Processign, *IEEE Transactions on Computers*, Vol. 25, pp. 1336–1346.

Goldberg, M., D. Goodenough, M. Alvo, and G. Karam, 1985. A Hierarchical Expert System for Updating Forestry Maps with Landsat Data, *IEEE Proc.*, Vol. 73, No. 6, pp. 1054–1063.

Goldberg, M., G. Karam, and M. Alvo, 1983. A Production Rule-Based Expert System for Interpreting Multi-Temporal Lansat Imagery, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, D.C., pp. 77–82.

Gonzalez, R., and M. Thomason, 1978. *Syntactic Pattern Recognition*, Addison Wesley, Reading, Mass., 282 p.

Gonzalez, R., and P. Wintz, 1977. *Digital Image Processing*, Addison Wesley, Reading, Mass., 431 p.

Goodenough, D., M. Goldberg, G. Plunkett, and J. Zelek, 1987. An Expert System for Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25, pp. 349–359.

Hanson, A., and E. Riseman, 1978. VISIONS: A Computer System for Interpreting Scenes. *Computer Vision Systems* (A. R. Hanson and E. M. Riseman, editors), Academic, New York, pp. 303–333.

Haralick, R., 1979. Statistical and Structural Approaches to Texture, *Proceeding of IEEE*, Vol. 67, No. 5, pp. 786–804.

———, 1985. Survey-Image Segmentation Techniques, *Computer Vision, Graphics, and Image Processing*, Vol. 29, pp. 100–132.

Harlow, C., 1973. Image Analysis and Graphs. *Computer Graphics and Image Processing*, Vol. 2, No. 1, pp. 60–72.

Harlow, C., S. Dwyer III, and G. Lodwick, 1976. On Radiographic Image Analysis, *Digital Picture Analysis* (A. Rosenfeld, editor), Springer-Verlag, Berlin, Chapter 3, pp. 65–150.

Harlow, C., M. Trivedi, and R. Conners, 1985. Use of Texture Operators in Segmentation, *Applications of Artificial Intelligence II*, SPIE, Arlington, Virginia, No. 548, pp. 10–18.

Harlow, C., M. Trivedi, R. Conners, and D. Phillips, 1986. Scene Analysis of High Resolution Aerial Scenes. *Optical Engineering*, Vol. 25, No. 3, pp. 345–355.

Harmon, P., and D. King. 1985. *Expert Systems: Artificial Intelligence in Business*, Wiley & Sons, New York.

Havens, W., and A. Mackworth, 1983. Representing Knowledge of the Visual World, *IEEE Computer*, Vol. 16, No. 10, pp. 90–96.

Hayes-Roth, F., D. Waterman, and D. Lenat, 1983. *Building Expert Systems*, Addison-Wesley, Reading, Mass.

Horn, B., 1978. *Vision Review*, M.I.T., Artificial Intelligence Laboratory, No. 157, Artificial Intelligence Working Paper, 17 p.

Hutchinson, C., 1982. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement, *Photogrammetric Engineering & Remote Sensing*, Vol. 48, pp. 123–130.

IntelligenceWare, 1987. *Intelligence/Compiler User's Manual*, Los Angeles, CA 90045.

Jensen, J., 1986. *Introductory Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Englewood Cliffs, N.J.

Kanade, T., 1977. Model Representations and Control Structures in Image Understanding, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 1074–1082.

Kanal, L., 1974. Patterns in Pattern Recognition: 1968–1974, *IEEE Transactions on Information Theory*, Vol. 20, pp. 697–722.

Kanal, L. and B. Chandrasekaran, 1972. On Linguistic, Statistical and Mixed Models for Pattern Recognition, *Frontiers of Pattern Recognition* (S. Watanabe, editor), Academic Press, New York, pp. 163–192.

Kanal, L., and A, Rosenfeld, (editors), 1981. *Progress in Pattern Recognition*, North-Holland, Amsterdam, 391 p.

Kapur, J., P. Sahoo, and A. Wong. 1985. A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram, *Computer Vision, Graphics and Image Processing*, Vol. 29, pp. 273–285.

Kettig, R., and D. Landgrebe, 1976. Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects, *IEEE Transactions of Geoscience Electronics*, GE-14, pp. 119–26.

Levine, M., and S. Shaheen, 1981. A Modular Computer Vision system for Picture Segmentation and Interpretation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-3, No. 5, pp. 540–556.

Li, R., and K. Fu, 1976. Tree System Approach for Landsat Data Interpretation, *IEEE Symposium on Machine Processing of Remotely Sensed Data*, Purdue University, West Lafayette, Indiana, pp. 2A-10–2A-16.

Lillesand, T., and R. Kiefer, 1987. *Remote Sensing and Image Processing*, John Wiley and Sons, New York.

Marr, D., 1982. *Vision*, W. H. Freeman and Company, San Francisco.

Matsuyama, T., 1987. Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing, *IEEE Transactions on Geoscience and Remote Sensing*, GE-25, No. 1, pp. 305–316.

Matsuyama, T., and V. Hwang, 1985. SIGMA: A Framework for Image Understanding of Bottom-Up and Top-Down Analysis. *Proc. 1985, Inter. Joint Conference on Artificial Intelligence*, pp. 1–18.

McKeown, D., 1987. The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems, *IEEE transactions on Geoscience and Remote Sensing*, GE25, No. 3, pp. 330–348.

McKeown, D., W. Harvey, and J. McDermott, 1985. Rule-Based Interpretation of Aerial Imagery, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7, No. 5, pp. 570–585.

Merchant, J., 1984. Using Spatial Logic in Classification of Landsat TM Data, Proc. *IEEE 1984 IX Pecora Symposium on Spatial Information Technologies for Remote Sensing Today and Tomorrow*, pp. 378–385.

Mero, L., and T. Vamos, 1981. Medium Level Vision, *Progress in Pattern Recognition* (L. Kanal and A. Rosenfeld, editors), North-Holland Publishing Company, Amsterdam, pp. 93–122.

Minsky, M., 1975. A Framework for Representing Knowledege, *The Psychology of Computer Vision* (P. Winston, editor), McGraw-Hill, New York.

Mintzer, O. W., 1989. Research in Terrain Knowledge Representation for Image Interpretation and Terrain Analysis, *U.S. Army Symposium on Artificial Intelligence Research for Exploitation of Battlefield Environment*, 1–16 Nov, 1989 El Paso, Texas, pp. 277–293.

Moik, J., 1980. *Digital Processing of Remotely Sensed Images*, NASA SP-431, U.S. Government Printing Office.

Mooneyhan, D., 1983. The Potential of Expert Systems for Remote Sensing. *Proceedings, 17th International Symposium on Remote Sensing of Environment*, San Francisco, California, pp. FP1 4.1–4.5. or pp. 1–14.

Morris, W. (editor), 1979. *American Heritage Dictionary*, Houghton Mifflin Co., Boston, Mass.

Nadler, M., 1982. Hybrid Pattern Recognition: A Synthesis of the Numerical and Qualitiative Approaches, *Pattern Recognition Theory and Applications* (J. Kittler, K. Fu, and L. Pau, editors), D. Reidel Publishing Company, Dordrecht, Holland, pp. 177–281.

Nagao, M., and T. Matsuyama, 1980. *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York, 199 p.

Narasimhan, R., and D. Argialas. 1989. Computational Approaches for Handling Uncertainties in Terrain Analysis. *Technical Paper, Annual Convention of American Society for Photogrammetry and Remote Sensing*, 2–7 April 1989, Baltimore, Maryland. Vol 3, pp. 302–310.

Narasimhan, R., 1974. The Role of Syntactic Models in Picture Processing, *Proceedings of IFIP Congress, Information Processing 74*, Stockholm, pp. 743–747.

Nazif, A., and M. Levine, 1984. Low Level Image Segmentation: An Expert System, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, No. 5, pp. 555–577.

Nevatia, R., 1982. *Machine Perception*, Prentice-Hall, Englewood Cliffs, N.J.

Niblack, W., 1986. *An Introduction to Digital Image Processing*, Prentice Hall, Englewood Cliffs, N.J., 215 p.

Nicolin, B., and R. Gabler, 1987. A Knowledge-Based System for the Analysis of Aerial Images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 25, No. 3, pp. 317–329.

Ohlander, R., K. Price, and D. R. Reddy, 1978. Picture Segmentation Using a Recursive Region Splitting Method, *Computer Graphics and Image Processing*, Vol. 2, pp. 313–333.

Pavlidis, T., 1977. *Structural Pattern Recognition*, Springer-Verlag, Berlin, 301 p.

———, 1982. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, Maryland, 416 p.

Peli, T., and D. Malah, 1982. A Study of Edge Detection Algorithms, *Computer Graphics and Image Processing*, Vol. 20, pp. 1–22.

Philipson, W., 1980. Problem Solving with Remote Sensing, *Photogrammetric Engineering & Remote Sensing*, Vol 46, No. 10, pp. 1335–1338.

———, 1986. Problem Solving with Remote Sensing: An Update, *Photogrammetric Engineering & Remote Sensing*, Vol. 52, No. 1, pp. 109–110.

Richards, J., D. Landgrebe, and P. Swain, 1982. A Means for Utilizing Ancillary Information in Multispectral Classification, *Remote Sensing of Environment*, Vol. 12, pp. 463–477.

Robinson, A., and B. Petchenik, 1976. *The Nature of Maps*, The University of Chicago Press, Chicago, 138 p.

Robinson, V., A. Frank, and M. Blaze, 1986. Expert Systems and Geographic Information Systems: Review and Prospects, *Jrnl. of Surveying Engineering*, Vol. 112, pp. 119–130.

Robinson, V., A. Frank, and H. Karimi, 1987. Expert System for Geographic Information Systems in Resource Management, *Applications in Natural Resource Management*, Vol. 1, pp. 47–57.

Rosenfeld, A., 1986. Expert Vision Systems: Some Issues, *Computer Vision, Graphics, and Image Processing*, Vol. 34, No. 1, pp. 99–117.

Rosenfeld, A., and A. Kak, 1982. *Digital Picture Processing*, Academic Press, New York, 453 p.

Rosenfeld, A., Hummel, R., and Zucker, S., 1976. Scene Labelling by Relaxation Operators, *IEEE Trans.*, SMC-6, No. 6, p. 420.

Samet, H., 1980. Region Representation: Quadtrees from Binary Arrays. *Computer Vision, Graphics and Image Processing*, Vol. 18, p. 88.

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

Schowengerdt, R., and H. Wang, 1989. A General Purpose Expert System for Image Processing, *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 9, pp. 1277–1284.

Shafer, G., 1976. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J.

Shapiro, L., 1979. Data Structures for Picture Processing: A Survey, *Computer Graphics and Image Processing*, Vol. 11, pp. 162–184.

———, 1980. A Structural Model of Shape, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, pp. 111–126.

———, 1985. The Role of AI in Computer Vision, *Proceedings of the Second Conference on Artificial Intelligence Applications*, IEEE Computer Society, Miami Beach, Florida, 11–13, Dec. 1985. pp. 76–81.

Suen, C., and R. Mori (editors), 1982. *Computer Analysis and Perception*, CRC Press, Boca Raton, Florida.

Swain, P., and S. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw-Hill.

Tailor, A., A. Cross, D. C. Hogg, and D. C. Mason., 1986. Knowledge-Based Interpretation of Remotely Sensed Images, *Image and Vision Computing*, Vol. 4, No. 2, pp. 67–83.

Thomason, M., and R. Gonzalez, 1981. Database Representation in Hierarchical Scene Analysis, *Progress in Pattern Recognition* (L. Kanal and A. Rosenfeld, editors), North-Holland, Amsterdam, Vol 1, pp. 57–91.

Tou, J., and R. Gonzalez, 1974. *Pattern Recognition Principles*, Addison-Wesley, Reading, Mass., 377 p.

Trivedi, M., and C. Harlow, 1985. Identification of Unique Objects in High Resolution Aerial Images, *Optical Engineering*, Vol. 24, No. 3, pp. 502–506.

Tsai, W., and K. Fu, 1980. Attributed Grammars—A Tool for Combining Syntactic and Statistical Approaches to Pattern Recognition, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 10, pp. 873–885.

Watanabe, S., 1971. Ungrammatical Grammar in Pattern Recognition, *Pattern Recognition*, Vol. 3, pp. 385–408.

Wharton, S., 1987. A Spectral-Knowledge-Based Approach for Urban and Land-Cover Discrimination, *IEEE Trans. on Geoscience and Remote Sensing*, GE-25, No. 3, pp. 272–282.

Winston, P., 1984. *Artificial Intelligence*, Second Edition, Addison Wesley.

You, K., and K. Fu, 1979. A Systematic Approach to Shape Recognition Using Attributed Grammars, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, pp. 334–345.

Zadeh, L., 1978. Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, Vol. 1., No. 1, pp. 3–28.

Zucker, S., 1976. Relaxation Labeling and the Reduction of Local Ambiguities, *Proceedings of the Third International Joint Conference on Pattern Recognition*, IEEE Press, pp. 852–861.

Zucker, S., A. Rosenfeld, and L. Davis, 1975. General Purpose Models: Expectation about the Unexpected, *International Joint Council on Artificial Intelligence, Fourth International Joint Conference on Artificial Intelligence*, Tbilisi, Georgia, USSR, pp. 716–721.

(Accepted 18 January 1990)