GODDARD GRANT
IN-82-CR
109770
66 P.

PENNSTATE

Interim Progress Report II

ADVANCED TECHNIQUES FOR THE STORAGE AND USE OF VERY LARGE, HETEROGENEOUS SPATIAL DATABASES

Principal Investigator:

Donna J. Peuquet

(NASA-CR-181517) ADVANCED TECHNIQUES FOR N88-12421
THE STORAGE AND USE OF VERY LARGE,
HETEROGENEOUS SPATIAL DATABASES. THE
REPRESENTATION OF GEOGRAPHIC KNOWLEDGE: Unclas
TOWARD A UNIVERSAL (Pennsylvania State Univ.) G3/82 0109770

submitted to
NASA/Goddard Space Flight Center

for the period
April 2, 1987 - June 31, 1987

By

The Pennsylvania State University
114 Kern Building
University Park, PA 16802

Grant No. NAG 5-798

Interim Progress Report II

This document represents the second interim progress report in the ongoing development of a prototype knowledge-based geographic information system in cooperation with NASA/GSFC personnel. The purpose of this overall project is to investigate and demonstrate the use of advanced methods in order to greatly improve the capabilities of GIS technology in handling very large, multi-source collections of spatial data in an efficient manner. The goal of this effort is to make these collections of data more accessible and usable for the earth scientist.

A proof-of-concept system, called KBGIS, was built at the University of California at Santa Barbara partially with NASA funding to investigate the use of new methods to improve the flexibility and overall performance of very large, multi-source, spatial databases. The system currently under construction at PSU, called GeoKnowledge, is based upon the design concepts and overall capabilities demonstrated in KBGIS and represents a continuation of that effort - toward a fully functional knowledge-based system.

The priority element of the current phase was the continuing refinement of techniques for efficient, non-exhaustive spatial search of a very-large, multi-source database. As detailed in the Mid-Year Progress Report, it was soon found that fundamental changes to the original demonstration system were needed before these refinements could be implemented. The conceptual characteristics of the relationships between spatial objects were examined to insure logical consistency and optimal efficiency within a highly flexible search facility.

A revised spatial knowledge representation and an elemental and consistent set of spatial relationships that operate on this representation are now fully functional within GeoKnowledge. A detailed conceptual description of the characteristics and use of the representational framework and associated spatial operators is attached. This description is currently being revised for publication in the technical literature.

Continuing research under Supplement No. 1 of the current grant will address previously postponed work in investigating the use of specialized AI tools and interactive graphics. More specifically, elements for continuing work during the period July 1, 1987 through June 31, 1988 include the following:

A. Investigation of specialized AI tools for use in spatial database applications
B. Begin development of a graphics interface
C. Continuing refinement of the heuristic spatial search facility

It is expected that the majority of this work will utilize the specialized knowledge engineering software on the Symbolics processor, while maintaining use of the MicroVax/VMS system as a backend database machine. Investigation of specialized AI tools will involve learning mechanisms and the development of consistency rules for the object database.

THE REPRESENTATION OF GEOGRAPHIC KNOWLEDGE: TOWARD A UNIVERSAL FRAMEWORK

Donna J. Peuquet
Department of Geography
The Pennsylvania State University
University Park, PA 16802

ABSTRACT

There is an urgent need to use geographic information systems (GIS) to manage extremely large databases containing data integrated from a number of imagery, cartographic and other sources for an increasing variety of applications. Current GIS technology has, however, revealed severe shortcomings in meeting these performance requirements.

The cause of this problem is that the spatial data models used in these systems have always been either hardware-driven, such as imagery data, or direct interpretations of the paper map. In both cases, a number of special characteristics of geographic data have not been taken into account. These characteristics include: First, natural geographic boundaries tend to be very convoluted and irregular. They consequently do not lend themselves to compact representation, and storage of these data can quickly become very large. Second, the data in digital form tend to be incomplete, imprecise and error-prone due to the complexity of the data and the characteristics of the data gathering process. Third, spatial relationships tend to be inexact or application-specific.

The present paper presents a new approach to building geographic data models that is based on the fundamental characteristics of the data represented. An overall theoretical

framework for representing geographic data is proposed. An example of utilizing this framework in a GIS context by combining Artificial Intelligence techniques with recent developments in spatial data processing techniques is then given.

INTRODUCTION

The primary bottleneck in the use of observational data in large-scale, real-world applications for many years was that data capture and database construction was a very slow and expensive process. As a direct result, operational databases tended to be limited in size, regardless of the intended scope of the completed database. With the advent of computerized spatial data handling systems, much attention was thus given to efficient data capture and input. Operational efficiency and flexibility were secondary considerations, at best. For both analytical and data management, representational schemes were developed on an ad-hoc basis using a heuristic approach, with little or no consideration of epistemological adequacy.

Due to recent advancements in automated data capture and input techniques and the subsequent availability of data from Landsat and other automated data capture devices, this situation has changed dramatically. There is now a rapidly expanding volume and variety of spatial data available in digital form. These data represent a very major investment and an extremely valuable resource. This rapid increase in data availability has caused a major crisis in the handling of these data.

Attempts to develop large-scale digital geographic databases and information systems has led to databases that are inefficient and inflexible. Such problems are also difficult to predict and are usually not discovered until after a substantial investment of time and money has been made. To make this situation worse, there is a rapidly increasing need for extremely flexible and efficient spatial databases that can be used as an analytical resource among a wide variety of applications and incorporate widely varying types of data. Examples include the current efforts by NASA an others to incorporate LANDSAT and other remote sensed imagery and cartographic data within the same database [Danielson, 1986]. Such efforts have served more to reveal the magnitude of the problem than to offer any immediate solutions.

The problem just described has two primary aspects: First, current techniques for conceptually representing and storing spatial data have exhibited severe limitations in the total volume of data that can be efficiently stored and quickly manipulated. Second, they are consistently limited in the range of types of information that can be easily represented.

The representation of geographic information is a central problem in the field of geography and in any field that studies phenomena on, over or under the surface of the earth. A representational scheme is required, and is in fact inextricably linked with the process of spatial analysis and the modelling of geographic phenomena. A representational scheme is also an integral part of the storage and subsequent use of geographic data in automated database and information systems. The validity of results of any analysis or model of a process can be quickly undermined if it is based on an inadequate or erroneous view of the geographic phenomenon under study.

The basic need is to be able to derive, with predictable results, a sufficiently precise and complete representation of the slice of reality involved for the application at hand. In order to do this, it is essential to develop new models or representational schemes for geographic data that are based on fundamental theory concerning the nature of geographic space. This need was recognized long ago [Berry, 1973; Lowenthal, 1961]. Recent developments in other fields have provided some tools and insights that can significantly aid in the development of such a theory.

The objectives of the current paper are therefore twofold; 1) to provide some insight into the long-term task of developing a fundamental theory and robust formalism for representing geographic space, and 2) help satisfy an immediate and practical need for efficient and flexible spatial data representation for all types of digital spatial database systems.

The remainder of the paper will be organized as follows:

First, techniques for modeling spatial phenomena and handling large heterogeneous data sets developed within several fields will be examined. Drawing on and combining these concepts, a set of general principles for representing geographic phenomena will be suggested and the derivation of a specific model based on these principles will be discussed.

Lacking the current existence of a structured body of spatial theory, the approach taken is empirical, and draws upon data modeling concepts initially developed within the fields of database management systems and computer vision to develop a suggested overall framework for representing geographic knowledge. The specific model discussed here is being implemented within a prototype knowledge-based geographic information system. This initial implementation will hopefully serve both purposes, advancing current operational data structuring techniques for geographic information systems and serving as an empirical tool for the study and improvement of our understanding of geographic phenomena within a formalized framework.

MODELS OF SPATIAL PHENOMENA

The complete enumeration of all observations and all possible relationships among these observations for all but very small data sets has proven impossible on a practical basis. The data must therefore be structured in a way that implies much more information than is explicitly stored. Consequently, any model of geographic space is of necessity imprecise and incomplete. If data stored utilizing such a model is to posses a known level of accuracy, the model must also incorporate a method of providing integrity and consistency checks. If the collection of data is also to be both large and efficient, there must also be a way of retrieving specific information without an exhaustive search.

Being able to structure data in such a way requires higher-level knowledge concerning the nature of the phenomenon represented and how component elements interact. It also requires techniques for representing and using that knowledge in a consistent and unified manner. Two fields that can provide insight into this problem are Computer Vision and Database Management Systems. Both have developed overall schemes for representing information as well as methodologies for implementing these schemes. Computer Vision deals with the spatial realm and has drawn heavily from cognitive theory. Database Management Systems has not paid particular attention to spatial problems, but has always emphasized techniques for handling very large volumes of data. Before examining aspects of these fields, some comments on cartographic models as representations of geographic space are appropriate.

CARTOGRAPHIC MODELS

The most universal and well-known representational scheme for geographic phenomena is the paper map. Every cartographic representation implies some view of the world, but these are not based on any formal theory of how to represent geographic space. The process of designing maps for the storage and retrieval of geographic information developed as a manual process that is more of an art than a science. The cartographer often takes liberties with reality in order to achieve a desired visual effect as well as to compensate for apparent irrational responses of the human eye-brain mechanism. Drawing a map, as well as retrieving information from one is thus an intuitive process and as such is not amenable to being cast into a structured universal framework or to being programmed into a computer. The necessary distinction between cartographic and digital representations of geographic space and a need for a unifying theoretical base for both has been recognized [Chrisman, 1977].

We will now look to developments in other fields for insight on how to characterize and formulate an overall conceptual framework and fundamental theory for representing geographic space.

SPATIAL MODELS IN COMPUTER VISION

Central to the field of Computer Vision is the development of efficient and robust models of space, and ultimately the representation and interpretation of spatial knowledge. This field has developed with two complementary problems; the practical problem of how to make computers ´see´ and the theoretical problem of developing a better understanding of how humans perceive the world. These are associated with the fields of Robotics and Cognitive Psychology, respectively.

The basic difficulty with robust models of space is that, as previously stated, there can never be a single model or view of the world that incorporates everything. Perceptions of the world vary among individuals and depend on the particular task at hand. An interior decorator´s view of a chair would likely be different than that of a structural engineer. Similarly, a geomorphologist´s view of a mountain would be different from a climatologist`s or a botanist`s, yet they would all recognize the same entity as a mountain. The views of individuals may also change over time. A mountain may look very different in summer than it does in winter, but would still be recognized as a mountain.

Noting these varying views, Gibson [1966] recognized that the key problem in understanding how humans perceive and model the world is in identifying the invariant or essential properties of the real world. This led to what Marr called the ´primal sketch´ [1982] He based the first unified theory for representing the seen world in an empirical context on this concept. He stated that such a representation should include some type of ´tokens´. These tokens represent individual entities that can be derived reliably from the image and can be assigned specific values for attributes, such as size or orientation. He then drew together the following physical assumptions regarding the overall spatial arrangement of these tokens as a universal and integral set. Each of these had been individually known within geography and other fields as fundamental characteristics of geographic space:

1.) Existence of surfaces - the visible world can be regarded as composed of continuous, smooth surfaces whose spatial structure may be elaborate,

2.) Hierarchical organization - the spatial organization of entities is often generated by a number of different processes, each operating at a different scale,

3.) Similarity - the items on a given surface responding to a process at a given scale tend to be more similar to one another in spatial organization, size and other attributes than to other items on that surface, 4.) Spatial continuity - spatial distributions generated on a surface by a single process tend to exhibit some sort of organized pattern,

5.) Continuity of discontinuities - spatial cohesiveness of entities and of spatial patterns results in a tendency toward smooth boundaries between them.

In general terms, his approach is to build up descriptive primitives, from the most detailed level up in almost a recursive manner, producing hierarchical groupings of entities and spatial patterns. This is an abstraction process where the tokens refer to increasingly abstract properties of the image at higher levels of the hierarchy. How to determine ´meaningful´ groupings can almost never be determined directly from the scene (i.e., the observed data). Some higher-level knowledge concerning the nature of the given phenomena involved must be employed. The higher-level knowledge or conceptual view is also organized and used differently from the ´raw image´ or seen view.

DATA MODELS FOR DATABASE MANAGEMENT SYSTEMS

In order to find a better approach for representing geographic information, we can also derive insight by studying current techniques initially developed within the field of Database

Management Systems (DBMS) for modeling non-spatial data related to business applications (e.g., payroll and inventory). Although the first use of computers for such applications began at approximately the same time as the first use of computers for geographic data, DBMS technology now seems to have progressed to a much more advanced state. Many studies have been done on how to apply the principles of state-of-the-art relational databases in an operational geographic context [Shapiro & Haralick, 1980; Van Roessel,1986].

Developments in this field were driven by a need for efficiency and flexibility in a practical, implementational context. A uniform framework was seen as the means of achieving this. The fundamental rationale in the initial development of the relational database concept was to provide a unified and consistent model for structuring the data with minimal redundancy. The most successful approach developed within DBMS to date is known as the Relational Database Model.

This model is based on the ´relation´. Each relation is simply a table containing a set of individual data entities or observations that are related in some manner. Each row in a relation contains attributes pertaining to an individual element. Each column contains values for a specific attribute for all elements. The relational model is directly derived from the mathematical concept of relations as properties of ordered sequences. For example, the expression $x + y = z$ defines a three-place relation for the set of natural numbers. Much elegance and power of the relational model is derived from one characteristic: Relationships between entities or groups of entities are not explicitly stored, but act as <u>operators</u> on the tables to produce derived relations. These are relations that provide users with their own views of the database.

Since relations are sets, the basic set operators of union, intersection and negation also hold and are used as a basis of operations that define how these relational sets can be combined in what are known as the relational algebra and the relational calculus. The manner in which the relational operators can be used is limited and controlled by a group of built-in rules known as integrity constraints. These integrity constraints function to avoid irrational combinations and to minimize data redundancy, and are based on mathematical principles regarding the properties of relations. These are summarized here for later reference. In the following notation, $x \; R \; y$ is to be interpreted as; x is in the relation R to y.

<u>Reflexive</u> - A relation is reflexive when, for any object x;

$$x \; R \; x$$

In other words, for any object x, the relation also holds for itself. A mathematical example is $x <= x$. This characteristic would hold for very few real-world examples.

Transitive - A relation is transitive if for all objects x, y and z;

if $x R y$ and $y R z$, then $x R z$

Examples of this relation include 'equal' and 'ancestor'.

Inverse - A relation R2 is called the inverse of relation R1 if;

$x R_1 y R_2 x$

In other words, the application of R_2 to the result of R_1 yields the original input value. Examples of inverse relational operators are employer/employee and parent/child.

Symmetric - A relation is symmetric when, for any object x;

$x R y$ implies $y R x$

In other words, the relation works in both directions with respect to any given pair of objects. This also means that any reflexive relation is its own inverse. Examples include 'spouse' and 'sibling'.

Several inherent shortcomings were soon discovered in this overall model. The two foremost of these were that actual implementations proved too slow for databases of any size and that this model is well-suited only for data with a regular, homogeneous structure. Extensions to the basic relational model were subsequently investigated with the use of semantic data modeling techniques [Codd, 1979].

A number of extensions suggested by Codd, consistent with Marr´s approach, were based on abstraction mechanisms for combining atomic entities, properties and associations into meaningful, higher-order units. Codd, however, grouped these into two types; generalization and aggregation. Precedence also is introduced as a successor mechanism. These extensions together allow a hierarchical or heterarchical data organization that is better suited to act as;

1.) a conceptual framework for representing a wide variety of data types, and
2.) a mediator between stored representations and user views.

To do this, the incorporation of the following additional representational forms into the data model were suggested;

1.) the inferential string-formulae provided by predicate logic for the representation of knowledge and application of inference techniques, and

2.) a labeled, directed hypergraph for higher-order relations and to support non-exhaustive search [Codd, 1979].

These extensions using techniques developed in the Artificial Intelligence community resulted from the observation that the relational calculus used in relational database management systems is precisely equivalent to the predicate calculus used for logic programming [Gallaire & Minker, 1978].

The use of this rule-based, graph-theoretic approach to represent inexact and view-dependant properties and relationships has proven to be much more suitable than fuzzy logic, as had been formerly proposed for such contexts [Zadeh, 1974].

The extended relational database model was employed by Meier and Ilg [1986] in a geographic context to handle spatial relationships. They proposed the graph grammar approach as a method of preserving consistency through arbitrary sequences of spatial operations. All consistent states are described by a structure graph and the transitions are given as sequences or rules.

This was demonstrated to potentially be a powerful mechanism for modeling spatial relationships as operators. Nevertheless, it was seen to be severely limited by a bewildering number and variation of potential spatial relationships and by a complex of often unpredictable side effects that can be produced by combining these relationships in arbitrary sequences.

The field of Database Management Systems, therefore, has provided a number of valuable concepts for a general model of geographic phenomena, although both geographic theory and direct use of the relational model in its current form are not adequate for this task. The problem of spatial relationships can only be handled by reducing the set of all spatial relationships into a small set of atomic or primitive spatial relationships with known characteristics. From this, formalized rules for combining operations and formulating higher-order relations can be derived systematically.

As a starting point for development of an overall framework for representing geographic phenomena, a robust definition of a data model that has evolved within this field can be employed. This definition can be summarized as follows:

> A data model may be defined as a general description of specific sets of entities and the relationships between those sets of entities. An entity is a thing which exists and is distinguishable; i.e., we can tell one entity from

> another. An entity set is a class of entities that possesses certain common characteristics [Ullman, 1982, pp 12-17].

Given this definition, a chair, a person and a mountain are each individual entities, whereas chairs, people, and mountains are each entity sets. Relationships include such things as ´left of´, ´taller than´ or ´parent of´. Both entities and relationships can have attributes, or properties. These associate a specific value from a domain of values for that attribute with each entity in an entity set. For example, a mountain may have attributes of size, elevation and geologic strata, among others.

A comparable definition of a data model was given by Codd [1981], who stated that a data model consists of three components; a collection of object types, a collection of operators and a collection of general integrity rules.

The formalized approach of DBMS technology will now be applied to the unified conceptual view of geographic space developed within computer vision in order to derive a more robust formalism for representing geographic space.

3. FORMALIZATION OF GEOGRAPHIC KNOWLEDGE

The very general problem of representing geographic phenomena can be broken down into several areas of investigation at several levels of abstraction. With the level of abstraction progressively increasing, these areas can be stated as;

1.) spatial information systems
2.) spatial knowledge structures and knowledge-based methodologies
3.) spatial understanding and formalization.

We need to begin at the highest level of abstraction, developing first a proposed conceptual framework of geographic space. The next step is to derive a knowledge structure from this framework that, in turn, can be implemented in an information system to empirically test the validity of the original model.

In the following section, therefore, a general conceptual framework will be derived. This will then be used as a ´canonical form´ for building a suggested spatial knowledge structure that is aimed toward real-world application. Immediate progression to this second step is viewed as a means of checking the model for robustness and completeness by translating the original model into a more detailed form.

BASIC COMPONENTS AND CHARACTERISTICS

The key to this overall process is to break down the phenomenon into its constituent parts and formally define each component and their interrelationships. Adopting the definition of a data model given in the previous section, it is assumed that a geographic data model can be considered to be composed of the following;

entities
properties
relationships.

Entities can be grouped into higher-order entities, and both entities and relationships have properties or attributes.

To summarize from the discussion so far, key characteristics of geographic phenomenon that need to be taken into consideration in formulating a representational framework for geographic phenomena are;

1.) the enumeration of entities, their properties and the relationships between entities tend to be imprecise, incomplete and view dependant, 2.) observed or recorded properties of entities can be numerous, and

3.) the boundaries of geographic objects tend to be convoluted and irregular.

Properties of entities can include general properties such as size, shape, color and height. They may also include domain-specific properties such as geologic strata in the case of mountains. With reference to a specific entity, each known property can be assigned a single value, a range of values, or a group of different values determined on differing measurement scales.

From these characteristics, the method of representation for spatial entities should;

1.) allow entities of any level of abstraction to be represented,

2.) use generalization, aggregation and successor functions as relational operators between entities and groups of entities, resulting in a conceptually hierarchical structure of entities,

3.) allow any number of attributes and more than one value for any attribute for any entity,

4.) allow for entities that may overlap

5.) allow for measurements at varying degrees of precision.

The hierarchical structure of entities would be defined through the use of abstraction functions as relational operators between entities and groups of entities. These operators would vary to suit the nature of the specific entities involved (i.e., they would need to be knowledge-based and domain-specific). Ultimately, this would constitute a taxonomy of geographic objects, such as the general example shown in Figure 1.

These five functional capabilities accommodate the first two of the three characteristics given above. They would serve to represent any type of knowledge, spatial or otherwise. It is in dealing with the third characteristic, the distinguishing spatial nature of geographic data, when problems arise. Vectors of x y coordinates defining the location and extent of individual entities could be stored as a property of each known entity. This can also be recorded multiple times, representing differing views at different scales or levels of precision.

Because each of these are not single values, their physical storage can represent a volume of data disproportionately larger than data stored for other properties. It also does not seem reasonable to assume that specifying scales or levels of precision are necessarily associated with specific application views. A botanist, geologist or any individual may choose a view at a highly generalized spatial scale or at a very detailed scale.

A much more important factor to be considered is the manner in which people acquire and use knowledge of the perceived world. All spatial questions can be classified into two basic categories that are logical duals of each other:

1.) Given a specific object or objects, what are its associated properties (one of these properties may be its location or locations)?

2.) What object or objects are present at a given location?

These correspond to object-based and location- or scene-based views, respectively. It is also noted that both of these questions can be generalized into a single form using the elemental components of a data model listed at the beginning of this section:

> Given a specific entity or group of entities, what are the values of their associated properties?

In the first question above, the spatial object is the entity and in the second question, the location semantically becomes the entity.

It is also possible to reverse the form of this generalized question to:

> Given a specific value for a specific property, what are the associated entities?
>
> e.g., Find the set of all mountains that are green.

This is assumed, however, to be a relatively unusual form of question.

These primary representation and usage characteristics of geographic information supports the use of a dual structure for modeling spatial phenomena and organizing spatial knowledge, one side being object-based and the other being location-based. This idea coincides with Marr´s overall framework of vision and spatial perception [Marr, 1982]. His approach to vision as being an information processing task also provides insight as to how the two sides of such a dual model would relate to each other. He asserts that processing of spatial information must begin with the raw scene; i.e., is initially location-based. Directly observed phenomena (e.g., reflectance values and discontinuities between them) must first be abstracted into selected, key characteristics of the scene, generating what he terms the ´primal sketch´. This sketch is interpreted using pre-existing knowledge, and objects are eventually associated with locations and groups of locations in the scene. Spatial objects are thus always derived as higher-order information.

The location-based representation should retain the same basic capabilities; i.e., allow for varying degrees of abstraction, use generalization and aggregation functions to define the values at varying levels, allow for any number of values for each location and allow for measurements at varying degrees of precision. Marr´s characteristics for low-level spatial information given in section 2.1, however, suggest perhaps a more regular structure than for the object-based representation.

Given a dual structure, it is helpful to slightly refine the definition of the elemental components of a spatial model to the following:

object-based representation	location-based representation
objects	locations
properties	properties
relationships	relationships

In this scheme, locations can also have properties or attributes, such as elevation, temperature, etc. These represent ´primitive´ properties, i.e., properties that are directly observable and are not necessarily characteristic of a particular object or objects. Relationships in a location-based context can take on a very special character - these are spatial relationships, such as ´contains´, or ´left-of´.

These concepts will now be cast into a more detailed, operationally-oriented structure.

REPRESENTATION OF SPATIAL ENTITIES

There has been much work recently in the field of Artificial Intelligence concerning the representation of knowledge pertaining to individual entities. These techniques have also been applied to spatial data [c.f., Tsotsos, 1984; Peuquet, 1984; Smith, Peuquet and Menon, 1987]. Central to these representational schemes is the expression of entity definitions in a formal language, such as first-order predicate calculus [Barr & Feigenbaum, 1981]. This approach allows the use of operators (e.g., and, or,not) in an expression to express a set of constraints that uniquely characterize that object. These are the properties that can be interpreted as the ´valid´ or essential properties of that particular object and may include size range, etc.

The set of all objects are implicitly arranged in interlinked hierarchies, as shown diagrammatically in Figure 1. These hierarchies are defined by the relationships to other objects contained within the object definitions. Such object relationships, for example, include ´is_a´ and ´component_of´.

In the location-based representation, locations are discretized into non-overlapping areal cells. Although space is perceived to be continuous, this is a necessary mechanism for recording variations over space in any formalized manner. For the sake of explanation and convenience, we divide our perceived universe in grid fashion into squares of uniform size. We can then logically superimpose increasingly coarser grids in hierarchical fashion to represent the same total area at increasing levels of generalization.

A convenient example of such a structure is the quadtree, as shown in Figure 2. This structure is based upon a recursive subdivision of a square area into four equal subunits. This results in a regular hierarchy of degree four and in cartographic terms produces a variable scale scheme based on powers of 2. This structure may not be the most appropriate for some types of information, but does provide a universally applicable, uniform structure that allows easy association of various types of information for the same areal unit. The quadtree also has been well studied and offers significant implementation advantages, as discussed in Peuquet [1984].

All locational properties can be logically viewed as individual surfaces layered on top of each other. All information pertaining to a single location at any level in the hierarchy (i.e., a node in the quadtree), however, should still be referenced with a single, unique locational index. Such indexing schemes have been discussed for quadtrees in Peuquet, Abel and Smith and others [Peuquet, 1984; Abel & Smith, 1983]. Each location contains information pertaining to each layer (i.e.,a single property for that location. For example; property value(s), as well as the name(s) of the specific method(s) used to abstract property values upward through the hierarchy. These methods are known as inheritance rules. This abstraction method may be specific to the particular property and may incorporate higher-level knowledge of the characteristics of that property. Information on how data for that layer are spatially distributed in the descendant, finer-resolution cells representing the same area would also be stored at individual locations throughout the hierarchy.

At the lowest level of the hierarchy, representing the finest locational resolution are the primitive, observed values. This is not necessarily at the same level in the hierarchy for all properties, in conformance with real-world observation.

RELATIONAL OPERATORS

As previously stated, there are two different types of relational operators in a spatial context;

abstraction relations, and
spatial relations.

Abstraction relations fall into two subtypes, one for combining geographic objects. We can call these taxonomic relations, and include ´is_a´ and ´component_of´. These operate on and define the object hierarchy, and they tend to be highly domain-specific. The other subtype combines the values of properties. These operators have a major function within the locational hierarchy, where the values of ´primitive´, observed properties are stored as discretized surfaces. These include average, mode, maximum, minimum, and any of a multitude of domain-specific aggregation or generalization techniques. Such techniques are well-studied and well-known. They also function on properties pertaining to objects, such as size and shape.

Spatial relations are unique to locational or spatial information. These relations are extremely important but not well-understood in any formal sense. Existing literature in this direction is very sparse and has primarily been done within the field of Computer Vision [Freeman, 1973; Winston, 1975; Evans, 1968; Haar, 1976; Claire, 1984]. In work to date, varying lists of ´basic´ spatial relations have been given. Freeman, for example, lists ´between´, ´touching´, left of´, ´right of´, ´above´ and ´below´ among a total of thirteen. Algorithmic models for these relations have been very simple and limited to the domain of regular geometric figures.

Since this seems to be a major missing element that is essential to the definition of any formalized representation of geographic knowledge, the remainder of this paper will focus on drawing together existing knowledge to try and provide some insights into this area in a geographic context, and examine potential gaps or flaws. A suggested framework for spatial relationships will be given that builds upon the overall spatial data model described thus far. Algorithmic approaches for specific relations will then be described.

On the basis of work performed by the author within an empirical context [Peuquet, 1984; Smith, Peuquet and Menon, 1987], it seems that all spatial relationships can be stated in terms of the following primitives;

boolean set operations
distance
direction

For example, the higher-order spatial relation ´nearest neighbor´ can be expressed as a series of relative distance relationships. Similarly, ´between´ can be expressed as a specific and limited combination of possible direction relationships. ´Touching´ or ´adjacent´ can be expressed as a special case of distance, where the distance between one object and a second object equals zero at one or more locations and is never less than zero. ´Left-of´, ´right of´, ´above´ and ´below´ are specific instances of the same relational concept (i.e., direction) in that the same model holds for all. A model for ´left of´ becomes a model for ´right of´ after performing a 180 degree coordinate rotation on the data.

This implies that developing an understanding of spatial relations in a formal, theoretical context is a much more tenable task than had been previously assumed, as only three spatial relationships, their characteristics and interactions need to be formally defined. All other spatial relations can then be defined in terms of these primitive relations and a set of combinatorial integrity rules. This is also particularly encouraging in the derivation of a complete and robust framework.

For the following discussion of spatial relationships, the quadtree model will be used as a basis for the algorithmic approaches given. Binary data layers will also be assumed for ease of exposition. The following conventions will be used for the following discussion:

1.) A black node denotes a quadrant at any level in the hierarchy that is homogeneous with respect to a particular data value. A white node denotes the absence of data (i.e., a null cell). A grey node denotes a cell that is not homogeneous with respect to a particular given data value.

2.) A grey node will always have at least one node below it in the hierarchy that is black. Black and white nodes will always be terminal nodes.

All algorithms described in the following sections operate by traversing the quadtree hierarchical structure.

Boolean Set Operations

Boolean set operators in the spatial domain are commonly known as map overlay operations. Conceptually, these are direct carry-overs from the non-spatial domain. All of the well-known algebraic and syntactic properties therefore apply [Behnke et. al., 1986]. The only distinguishing factor here is that the two sets represent sets of locations in space. If the two locational sets define two respective contiguous areas, then these operations are literal interpretations of the classical boolean diagrams, as shown in Figure 3a. As spatial set operations they do not, however, need to be single, spatially contiguous features (cf. Figure 3b).

Such operations on spatial data represented in tessellar form, usually a matrix of square cells, are fairly simple and straightforward. Algorithmic approaches for these boolean operations on quadtrees were presented by Schneier [1980]. These are special cases of the superimposition algorithms of Hunter and Steiglitz [1979].

Using the definitions for individual quadtree node ´colors´ given, it is seen that these same algorithms are applicable to multi-valued input data layers. The resultant layer, however, remains binary because of the nature of the process involved. This means that for the resultant data layer, a black cell is interpreted merely as an ´on´ cell and a white cell is interpreted as null or ´off´.

Intersection -

The quadtree set intersection algorithm of Schneier involves traversing two tree layers in parallel from the top down and selecting the appropriate action for one of only three conditions wherever the traversal reaches a black node: If a black node is encountered in both layers, then the corresponding node in the resultant tree layer is also black. If one layer is black and one is white, then the node in the resultant tree layer will be white. If one layer is black and the other is grey, then the corresponding node in the resultant tree layer is grey and the structure (i.e., the node colors) of the entire subtree below that node for that layer is also copied to the resultant tree layer. If the color encountered in both input layers are grey for a given node, then the descendant nodes are examined, recursively.

Union -

The quadtree union algorithm is very similar to the intersection algorithm. Again, both input layers are traversed in parallel from the top down. If a black node is encountered in either of the input layers, the color at the same node for the resultant layer is black. If one layer is white and the other layer at the same node is grey, the color in the resultant layer is grey. The structure of the entire subtree below that node is also copied to the resultant layer. Finally, if both layers encountered at a node are grey, then the entire process is repeated recursively for the descendant nodes.

Containment -

Containment (i.e., subset) is normally viewed as a binary, predicate relation in that it has a true/false value for any two given areas. In again traversing the two layers in parallel, assuming the test is to see if A is contained in B, the tree is descended breadth-first until a black node for A (i.e., a location completely covered by B) is reached. The color for B at the same node is then checked. If it is not also black, the value of the relation is ´false´ and the operation ceases. If both are black, then the descent of the tree continues until the next black node for A is found and the test repeats. If no more black nodes for A are present, the operation ceases and the relation is ´true´.

Summary -

It can be seen from the above that dealing with boolean spatial relationships is clearly defined and straightforward in a hierarchical implementation. The overall approach is also non-exhaustive by taking advantage of generalized information at higher levels in the hierarchy. Top-down traversal of the hierarchy also allows the entire process to be terminated at any selected higher level, yielding an approximate result at a chosen level of resolution.

Distance

Distance and direction relational operators are unique to spatial data. They are binary, as opposed to set, operators. The result of these two relational operators is also a single value, and not another set of locations. Given their binary nature, both distance and direction can be expressed in human terms in reference to either the locational or the entity domain. For example:

the distance between 41^{d} 30´ N, 81^{d} 30´W and 41^{d} N, 79´W

or

the distance between New York City and Cleveland.

However, since spatial relational operators operate only in the spatial domain by definition, entities must be translated into their locational descriptions. This brings up the question of scale. At some small scale, i.e., high in the locational hierarchy with low resolution representation, the locational description of an entity is represented as a single point location. At greater resolution, the entity may be perceived as a linear or areal feature in space and is represented as a set of locations. Polygonal and linear features cause complications for both distance and direction, as will now be described in the case of distance.

Distance between two point locations is clearly understood and is normally expressed in terms of one of three metrics, as follows. Let (x_1,y_1) and (x_2,y_2) be two points in cartesian coordinate space. Then,

1.) $E\ [(x_1,y_1),(x_2,y_2)] = (x_1-x_2)^2 + (y_1-y_2)^2$:euclidean

2.) $d_4\ [(x_1,y_1),(x_2,y_2)] = |x_2-x_1| + |y_2-y_1|$:city block

3.) $d_8\ [(x_1,y_1),(x_2,y_2)] = \max(|x_1-x_2|,|y_1-y_2|)$:chessboard

Distance is thus mathematically defined from point to point and is symmetric (i.e., a D b implies b D a). The problem arises with polygonal and linear spatial features on how to determine these two points. Is the distance between two linear or polygonal features defined as the minimum distance, the maximum distance or the distance between their centers of gravity or centroids? For all of these, the shape, sinuosity and relative positions of the two features can affect how distance can be determined algorithmically.

For geographic features, we must assume that the features can be areal, linear or point in nature, and have any arbitrary shape in any orientation in relation to each other. They can also be convex, concave and of arbitrary sinuosity. Polygons can be multiply connected (i.e., have holes). To outline a specific example of how distance can be determined for geographic features represented in quadtree form, an algorithmic approach for determining <u>minimum</u> distance between any two features will be briefly described. For a more detailed description, the reader is referred to Peuquet [1987b]. This approach takes advantage of the manner in which quadtrees hierarchically subdivide space. The basic steps are as follows:

1.) Find the smallest common quadrant that completely encloses both features.
2.) Recursively subdivide the quadrant until the features or portions of the two features, occur in separate quadrants. This will result in two or more pairs of adjacent quadrants at different levels of the hierarchy (cf., Figure 4). Quadrants containing parts of either feature that cannot be paired in this manner are discarded.
3.) For each quadrant in each pair, use ´line of sight´ relative to the adjacent quadrant to determine the approximate facing sides of the respective feature boundaries.
4.) Calculate minimum distance between the two ´visible´

sides for each pair of quadrants and take the minimum distance among all pairs.

Similar to the boolean operators, the distance algorithm given below can be used to generate approximate results by limiting the depth in the hierarchy used by the algorithm for both quadrant subdivision and calculation of visible boundaries.

Direction

By far the most complex spatial relational operator is direction (ie., relative position between two locational features). This relational operator is binary and, assuming a finite number of discretized directions, each specific direction is coupled with an inverse [Freeman, 1973]. For example;

a NORTH b   implies   b SOUTH a

Similar to direction, we can only specify an exact, quantitative value for this operator between two points. For the measurement of relative direction with respect to linear or areal features, we again have the problem of; between what two points on the two features, respectively, is the relation determined? The human response in this case is to simply be less precise. In other words, we use a generalized measurement such as north,

north-west, south, etc., instead of degrees of inclination from the horizontal. This is also all the precision we may wish to record for the direct relation between two points. Therefore, as we go up the locational hierarchy, there would be increasing tendency to use approximate directional measurements.

The approximate directional relationship between two polygons (e.g., left, above, beside, east, north), because it is approximate, is often dependent on human interpretation. The problem is made even more complex in the case of arbitrarily--shaped, non-point features because of the effects that relative size, distance, shape and orientation have on the perceived directional relationship. The rigidness of the interpretation can also be influenced by the application. A model that can handle all possible cases is consequently difficult to derive except in a very generalized form.

A number of researchers have offered insights into the perceptual characteristics of this and other spatial relations, most notably Freeman [1973], Winston [1975], Evans [1968] and Haar [1976]. Their models of the directional relationship, limited primarily to points and squares, have recently been integrated and extended to arbitrarily-shaped features [Peuquet, 1986].

The primary perceptual characteristic of generalized direction is that the area of acceptance for any given direction increases with distance. This implies that, in general, any procedural definition for this relational operator must incorporate a triangular geometry, as shown in Figure 5. A simple method for determining approximate direction in the quadtree locational representation is outlined below. This method incorporates the triangular geometry for more precise determinations and calculates the result relative to the two centers of gravity for the two features:

1.) Find the smallest quadrant that completely encloses feature A and also the smallest quadrant that completely encloses feature B.

2.) Adjust the relation so that it is in relation to the larger feature (and larger quadrant).

3.) If only a very general approximation is desired, divide the area around the larger feature into 8 possible directions according to the top, bottom and sides of the larger quadrant (cf., Figure 6) and stop.

4.) Otherwise, find all quadrants completely covered by feature B and the same for feature A. (In other words, find the complete spatial definition of each feature)

5.) Calculate the center of gravity for each feature.

6.) Calculate degrees of arc from the reference feature to the second feature relative to the two points. If exact measurement is desired, stop.

7.) Otherwise, from the center of gravity of the reference feature, assume the surrounding area outside of the feature is divided equally into eight possible directions defined as ranges in degrees of arc (cf., Figure 7).

This simple procedure as described may give erroneous results for intertwined features. For a more complete procedure that takes such situations into account, see Peuquet [1987a].

Unresolved Problems

The short discussion for each spatial relational operator above shows that the further development and understanding of such operators holds promise;

1.) by virtue of the small number of primitive relational operators, and

2.) because some understanding and adequate algorithmic approaches for primitives already exists.

It is quickly seen that there is a wide variation in how certain aspects of these primitives may be defined. Further verification that the three operators given in the present paper do in fact comprise the set of primitive spatial relational operators needs to be undertaken.

Distance and direction are normally defined as binary operators. Models developed for these relationships, and subsequently algorithms derived from these models, by definition assume the presence of only two features. This is often not the situation on how a human may pose a spatial question. For example, a typical question may be; "Find the locations of all nuclear power plants within 50 miles and upwind of any urbanized area within the U.S." Here, what is implied is a set operator that compares the set of all nuclear power plants with the set of all urbanized areas. An area for further research is therefore how to extend our current binary models of primitive spatial relational operators so that they can be effectively applied to <u>sets</u> of spatial features.

While this would increase the level of correspondence between the definition of spatial relational operators to human perception, there is perhaps a more important aspect. The definition of <u>all</u> spatial relational operators as set operators would allow the uniform application of set theory. This would significantly increase the potential power of any spatial relational algebra or relational calculus in defining how these operators can be combined in a formalized, mathematical sense.

There is obviously a very fuzzy line between objective and subjective definitions for idealized geometric relational definitions. This was quickly seen in the definition of direction in a necessary generalized form. Some influence of subjective or interpretive meaning is unavoidable by the very nature of the spatial model.

This brings up an obvious issue that has not been explicitly stated so far in the present discussion: There are wide variations in semantic meanings of spatial relations in natural language expressions. This is significantly beyond the scope of the current research. The first task is certainly to derive canonical geometric description functions for primitives and a mechanism for combining them in a strict, formalized manner. With this in hand, the problem of defining semantic deviations in context from these ´ideal´ forms, including definition of approximations, could be more easily handled. Past research in this area has so far revealed more problems than answers [Herskovits, 1985]. The eventual derivation of at least some general usage and integrity rules for combining spatial operators in varying contexts as well as flexible orderings would significantly enhance the overall power of the spatial model.

SUMMARY AND FUTURE DIRECTIONS

The elements and characteristics of a formalized conceptual framework has been discussed and an example of a structure for representing spatial knowledge has been described. From this it seems that the overall characteristics suggested (e.g., hierarchical structure, separation of locational and conceptual views and the ability to store knowledge at variable levels of completeness and precision), draws great support on the basis of an agreement of findings among related disciplines. Given a significant amount of research in the recent past, powerful methods for appropriately representing both locational and object views conforming to these characteristics are shown to be available.

This discussion, however, hints at many other issues. Several issues, unique to the geographic context, remain as major obstacles in using this as a functional knowledge representation for practical applications and prime areas for further theoretical research. The first, mentioned in the present paper, is in refining the definitions and understanding of primitive spatial relationships and how they interact so that, at minimum, a relational inference structure can be developed. This is needed before these primitives can be stated as formal definitions of higher-order relations and before integrity rules for combining operations can be defined.

The other is to further examine the functional linkages between the locational and object entity representations. In the discussion of the representation of spatial entities, it was mentioned that a ´locational indicator´ can be stored with the representation of any given object entity. This is the link between the locational and object views, and is therefore a critical component of the dual representation scheme suggested. But what form should this take? - Certainly not a complete locational definition in all but perhaps a very few cases, if ever. From a perceptual standpoint, this would be extremely rare for anyone to know the explicit coordinate definition for any spatial object. From an operational database standpoint, that would be redundant data already being stored in the locational representation. Point indices representing a centroid or center of gravity also do not make sense in either a logical or practical context.

There is currently a significant amount of research being conducted on the handling of large, heterogeneous data sets in geography as well as other fields that deal with both spatial and non-spatial data. It has been shown that much of what has been learned in the contexts of these other fields can be applied to solving the data handling problems within the geographic context, as well as to expand the theoretical foundation of Geography as a whole. It has also been shown that a unified framework for modeling geographic phenomena need not be as complex as had been previously anticipated. A suggested general direction has been given in this paper.

REFERENCES

Abel, D. & J. Smith, 1983. "A Data Structure and Algorithm Based on a Linear Key for a Rectangle Retrieval Problem."Computer Vision, Graphics, and Image Processing. Vol. 24, pp. 4-14.

Barr, A. & E. Feigenbaum (eds.), 1981. The Handbook of Artificial Intelligence, Vol II. Los Altos, Ca.: Wm. Kaufman.

Behnke, H. et. al. (eds.), 1986. Fundamentals of Mathematics. Vol. I. Cambridge, Mass: The MIT Press.

Berry, B., 1973. "A Paradigm for Modern Geography." in R. Chorley (ed.), Directions in Geography. London: Methuen.

Claire, R., 1984. "Algorithm Development for Spatial Operators." Proceedings, Pecora 9 Conference, Sioux Falls, S.D., pp. 213-221.

Codd, E., 1979. "Extending the Database Relational Model to Capture More Meaning." ACM Transactions on Database Systems. Vol. 4, No. 4, pp. 397-434.

Codd, E., 1981. "The Capabilities or Relational Database Management Systems." IBM Research Report RJ 3132.

Chrisman, N., 1977. "Concepts of Space as a Guide to Cartographic Data Structures," Proceedings, Harvard Advanced Study Symposium on Topological Data Structures. Cambridge, Mass.

Danielson, G. (ed.), 1986. The Planetary Data Systems Concept. Report prepared for NASA Solar Systems Exploration Division by the Planetary Data Systems Group Committee.

Evans, T., 1968. "A Heuristic Program to Solve Geometric Analogy Problems." in Semantic Information Processing, M. Minsky (ed.) Cambridge: The MIT Press.

Freeman, J., 1973. "The Modeling of Spatial Relations." Computer Graphics and Image Processing. Vol. 4., pp 156-171.

Gallaire, H. & J. Minker, 1978. Logic and Databases. New York: Plenum Press.

Gibson, J., 1966. The Senses considered as Perceptual Systems. Boston: Houghton Mifflin.

Haar, R., 1976. "Computational Models of Spatial Relations." technical report TR-478 MCS-72-0361-10, Dept. of Computer

Science, University of Maryland.

Herskovits, A., 1985. "Semantics and Pragmatics of Locative Expressions." Cognitive Science, Vol 9., pp 341-378.

Hunter, G. and K. Steiglitz, 1979. "Operations on Images Using Quadtrees." IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 1, pp. 145-153.

Lowenthal, D., 1961. "Geography Experience and Imagination Towards a Geographical Epistemology." Annals AAG, Vol 51, pp. 241-260.

Marr, D., 1982. Vision. New York: W.H. Freeman Co.

Meier, A. and M. Ilg, 1986. "Consistent Operations on a Spatial Data Structure." IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 8, No. 4, 532-538.

Peuquet, D., 1984, "Data Structures for a Knowledge-Based Geographic Information System." Proceedings of the International Symposium on Spatial Data Handling. Zurich, pp. 372-391.

Peuquet, D., 1987a. "An Algorithm to Determine the Directional Relationship between Arbitrarily-Shaped Polygons in the

Plane." Pattern Recognition, forthcoming.

Peuquet, D., 1987b. "An Efficient Distance Operator for Quadtrees." unpublished report, Dept of Geography, The Pennsylvania State University.

Russell, D., 1979. "An Open Letter on the Dematerialization of the Geographic Object." in: S. Gale and Gl Olsson (eds.), Philosophy in Geography, pp. 329-344. Dordrecht, Holland: D. Reidel Publishing Co.

Shapiro, L. and R. Haralick, 1980. "A Spatial Data Structure." Geoprocessing. Vol. 1, pp. 313-337.

Shneier, M., 1981. "Calculations of Geometric Properties Using Quadtrees." Computer Graphics and Image Processing, Vol. 17, pp. 296-304.

Smith, T., D. Peuquet and S. Menon, 1987. "KBGIS II: A Knowledge-Based Geographic Information System," International Journal of Geographical Information Systems. forthcoming.

Tsotsos, J., 1984. "Knowledge and the Visual Process: Content, Form and Use," Pattern Recognition, Vol 17, pp 13-27.

Ullman, J., 1982. Principles of Database Systems. Rockville, Md.: Computer Science Press.

Van Roessel, J., 1986. "Design of a Spatial Data Structure Using the Relational Normal Forms." Proceedings, Second International Symposium on Spatial Data Handling. Zurich, pp. 251-272.

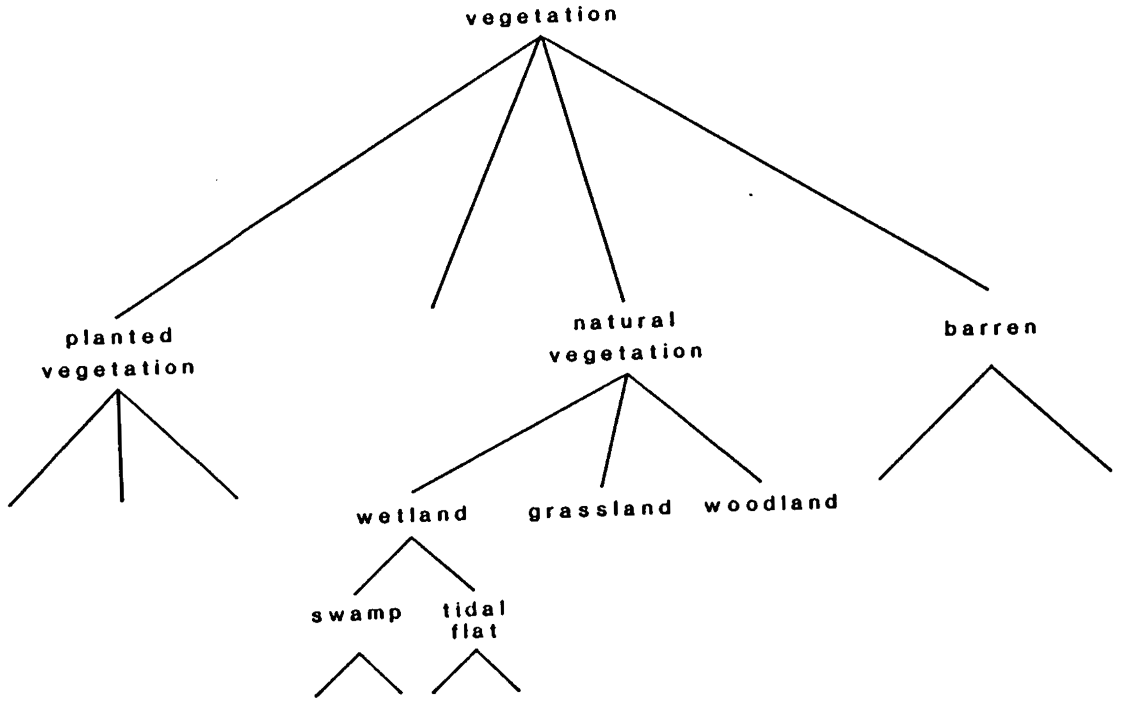
Winston, P., 1975. "Learning Structural Descriptions from Examples." in The Psychology of Computer Vision, P. Winston (ed.) New York: McGraw Hill.

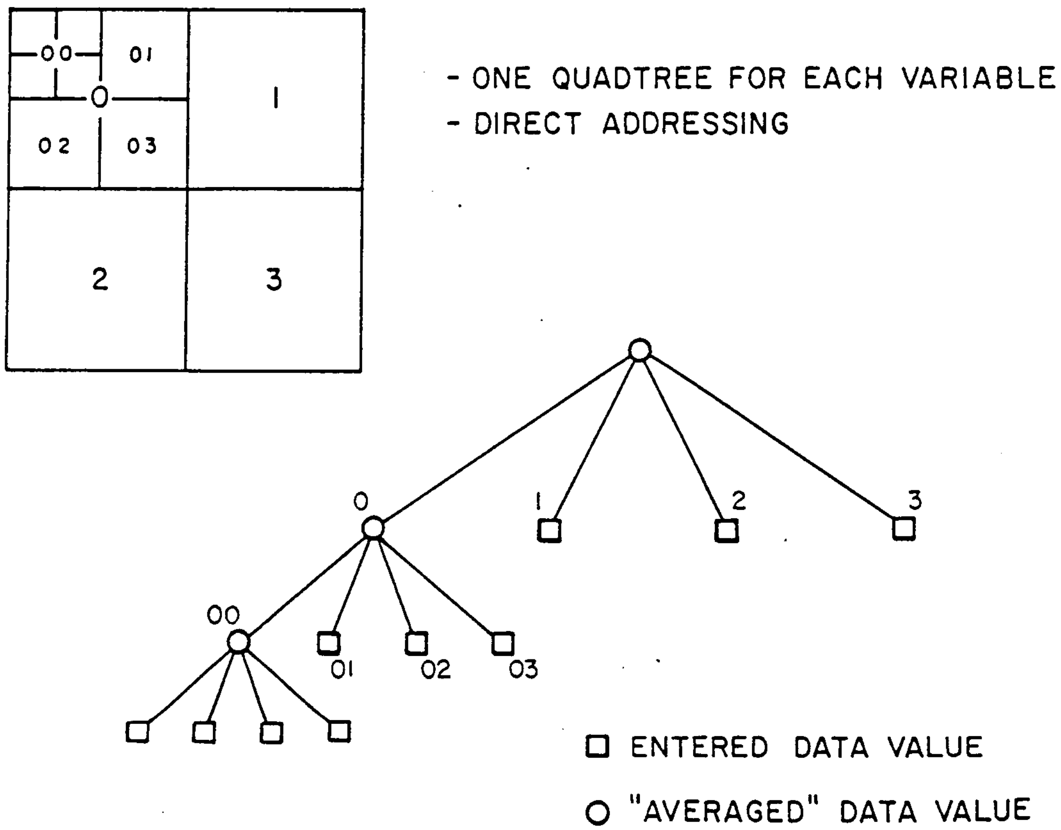
Zadeh, L., 1974. "A Fuzzy-algorithmic Approach to the Definition of Complex or Imprecise Concepts (Memo No. ERL-M474) Electronic Research Laboratory, Univ. of California, Berkeley.

LIST OF FIGURES

larger feature.

7 A more precise determination of relative position can be caluclated by radiating sectors from the centroid of the larger feature as boundaries between descretized directions.

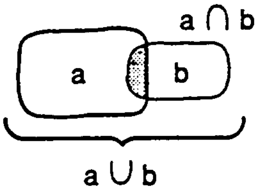

vegetation
planted vegetation
natural vegetation
barren
wetland
grassland
woodland
swamp
tidal flat



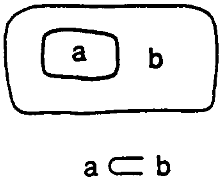
- ONE QUADTREE FOR EACH VARIABLE
- DIRECT ADDRESSING

□ ENTERED DATA VALUE

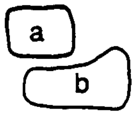
○ "AVERAGED" DATA VALUE



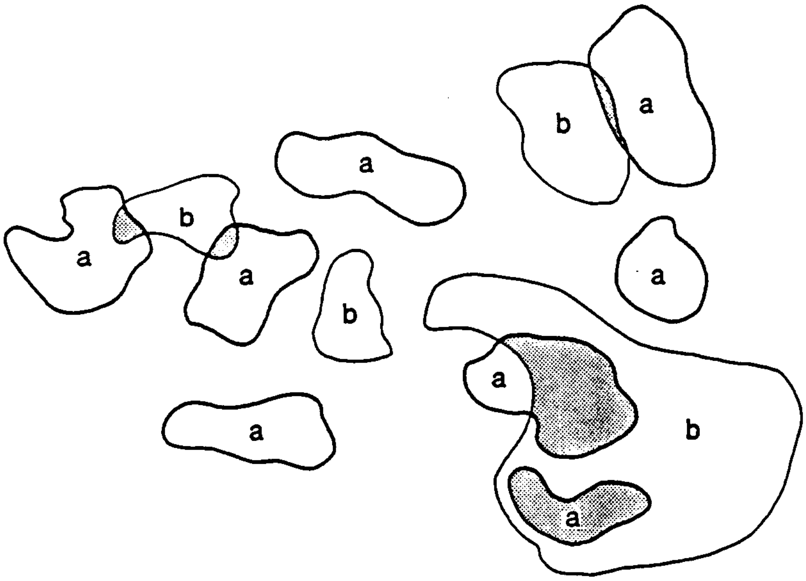
intersection & union

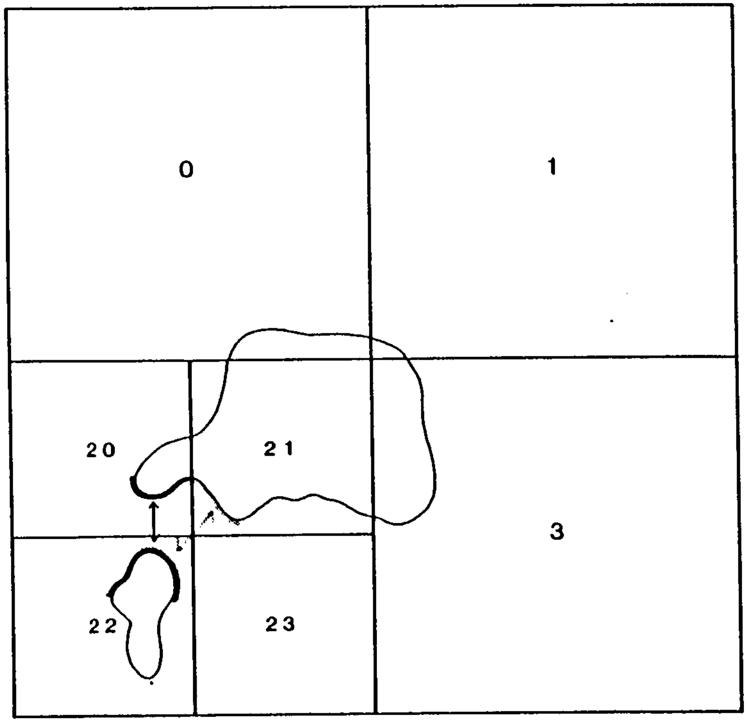


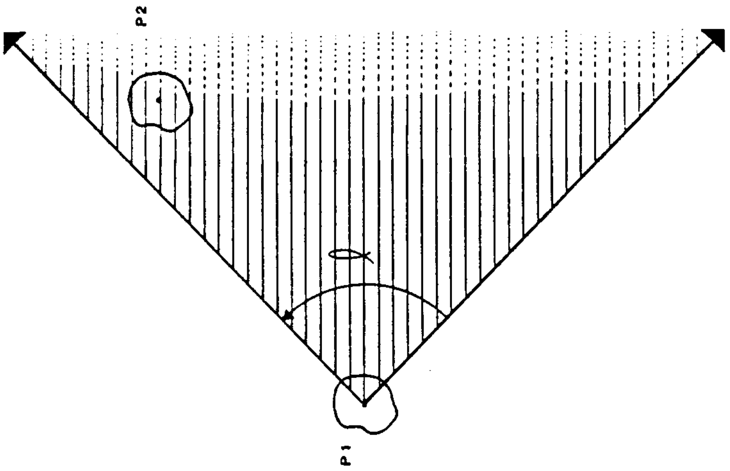
containment

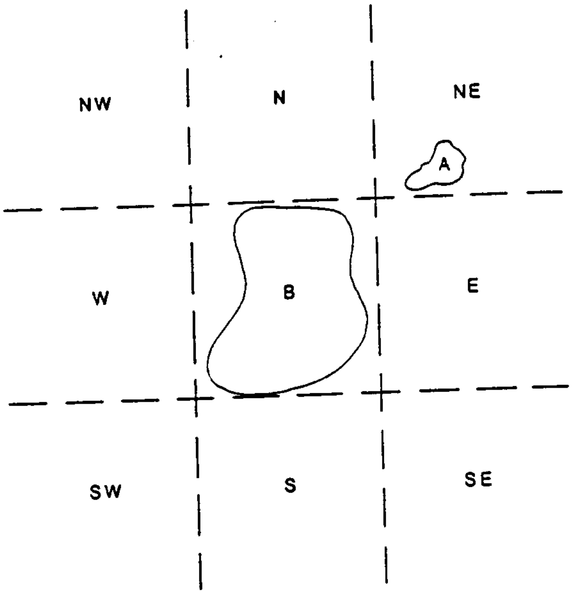


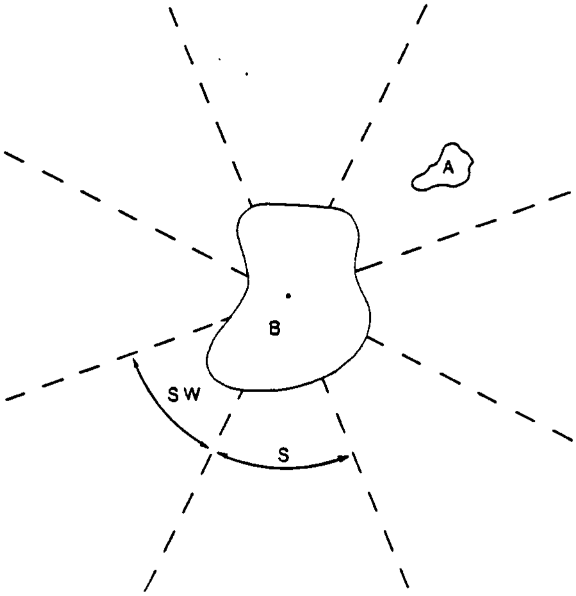
a and b disjoint




0
1
20
21
3
22
23


P2
θ
P1


NW
N
NE
A
W
B
E
SW
S
SE


A
B
SW
S