A-B19

Bruegger, B. P., and A. U. Frank. "Hierachical Extensions of Topological Datastructures." Paper presented at the 19th Congress of International Federation of Surveyors, Helsinki, Finland, 10.-19. June 1990 1990.

Bud P. Bruegger
Andrew U. Frank
National Center for Geographic Information and Analysis
and
Department of Surveying Engineering
University of Maine
Orono, ME 04469
BUD@MECAN1.BITNET
FRANK@MECAN1.BITNET

Baltimore 89

ABSTRACT

Topological data structures have become an accepted method for modelling spatial objects, with cells being the fundamental units of which objects are composed. Known topological data structures offer only a single degree of spatial resolution for modelling objects. Objects can therefore not easily be displayed in different resolutions. Complex objects are typically composed of a large number of cells and their representation requires a large amount of data. This causes their handling to be slow and inefficient. Typically, in multipurpose geographic information systems (GIS) small spatial objects coexist with very large objects. The complex needs of different user groups requires the display of spatial objects in several resolutions. Topological data structures are necessary, but not sufficient by themselves, for the implementation of multipurpose GIS.Additional structures are needed to obtain the considerable cost savings of data sharing.

To overcome these problems, a formal, dimension-independent approach for building multiple, hierarchically related representations of spatial objects is proposed. The representation is formed of layers of topological data structures that provide different spatial resolution and amount of detail. The layers are interconnected by hierarchical relations between their cells. The benefits of the structure are that objects can be accessed in different degrees of detail. Small and large objects can be handled efficiently and output of adequate resolution can be produced directly. This makes this multiple topological representation well suited for the implementation of multipurpose GIS.

INTRODUCTION

Topological data structures have become an accepted method for representing spatial objects [White 1985]. One such structure, the cell complex, shows promise of high utility: discussions of cell complex methods may be found in [Frank 1986] and recent implementation examples may be found in TIGER [Kinnear 1988] [Boudriault 1988] and in the TIGRIS system [Herring 1988].

They explicitly model the basic topological relationships between all spatial objects and are based on Corbett's topological model of the map [Corbett 1979]. Topological

* This research was partially funded by a grant from NSF under No. IST-8609123 (principal investigator: Andrew Frank). The support from NSF for the NCGIA under grant No. SES 88-10917 is gratefully acknowledged.

structures are designed to organize objects of a single degree of spatial resolution and detail.

Spatial objects are organized as aggregations of cells--the fundamental units modelled in the structure. As will be shown later, objects that are relatively large compared to the spatial resolution chosen for the structure and compared to the size of other spatial objects typically consist of a large number of cells. This means that it takes a relatively large amount of data to represent them in the structure.

Problems arise when dealing with large objects which are composed of many smaler ones. Handling large amounts of data in computers is time consuming. For topological data structures this means, that large spatial objects are handled much less efficiently than small objects. The system performance can become unacceptable when working with large objects, while no problems occur with small objects.

These problems observed with topological data structures are a major impediment for the implementation of multipurpose geographic information systems (GIS). Typically, in multipurpose GIS the range of sizes for spatial objects is very large. Because of different needs of different user groups, very large objects such as nations and continents coexist with very small ones such as parcels. The performance of such a GIS is not acceptable for the user groups dealing with objects in a large scale view. This considerably limits the number of purposes satisfied by a single GIS and, therefore, limits the cost savings possible through sharing data acquisition and updates among several users. A second problem is the single spatial resolution offered by the topological data structure. Different user groups require different degrees of resolution for the view of their objects. A single topological data structure cannot satisfy these needs.

In order to overcome these problems of single topological representations, structures have to be found that provide several representations of a spatial object [Bruegger 1989]. The representations differ in the amount of detail and thus detail describing the spatial objects. This paper proposes a multiple representation of varying spatial resolution. Figure 1 sketches the structure of this multiple representation. Several layers of topological data structures form the representation. The same object is represented in several layers in different degree of detail. The layers are interconnected by hierarchical relations between their cells.

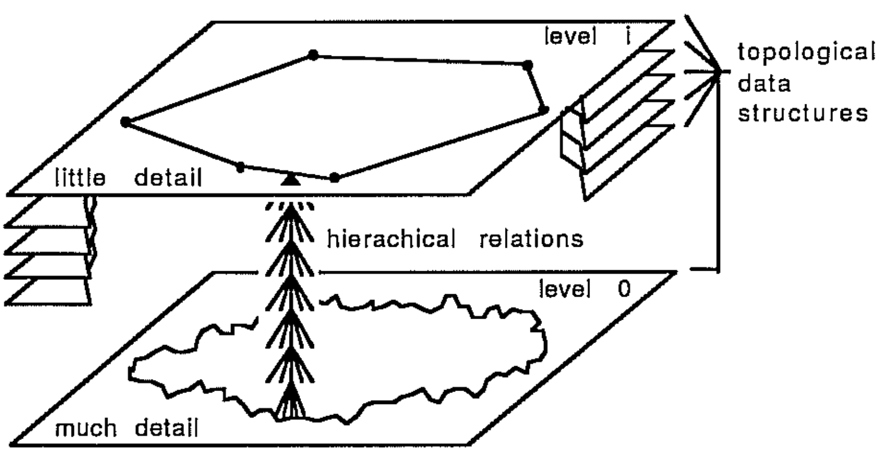

Figure 1: Structure of the multiple representation proposed in this paper.

In the field of spatial data handling, hierarchical structures have often been proposed as a means to relate multiple representations of different spatial resolution. Strip trees [Ballard 1981], Quad- and Octrees [Samet 1984] are typical examples. This research integrates their concepts and adapts them to topological data structures. The structure proposed in this paper can be considered a generalization of strip trees. Jones and Abraham propose a similar hierarchical structure which is not based on topological data structures [Jones 1986].

This research on hierarchies over topological data structures is part of a more extensive investigation of a spatial theory [Frank 1987b]. Earlier, a prototype using cell complexes to organize spatial objects has been implemented [Jackson 1989] [Pullar 1988]. Experience with the prototype demonstrated the problems of single topological representations.

This paper is organized as follows: In the introduction, problems encountered with single topological representations are described. Chapter 2 explains the concept of topological data structures and investigates the reasons for the problems described in the introduction. In chapter 3, a special case of a multiple representation is proposed to solve the problems. The conclusions stress the importance of the proposed solution for the implementation of multipurpose GIS and describes the planned continuation of this research.

TOPOLOGICAL DATA STRUCTURES

Spatial Properties

Topological data structures describe the geometry of spatial objects in terms of 0-cells (nodes), 1-cells (edges), 2-cells (faces), and 3-cells (volumes). The topological relations between cells are explicitly modelled in the data structure and build the bases for the extraction of topological information. In contrast, non- topological approaches derive these relations from metrics, i.e. coordinates.

No intersections of cells are allowed in topological data structures, i.e., the intersections have to be introduced as cells of their own, and the formerly intersecting cells have to be split up into several cells [Frank 1986]. Thus, all the possible intersections of the cells have to be determined as the new cells are introduced into the structure [Frank 1987a]. Figure 2 shows an example of two intersecting 2-cells. Cells introduced because of the intersection are highlighted. The two intersection points of their boundaries are inserted as new nodes, the intersecting 1-cells of the 2-cell-boundary are split, the intersection area is introduced as a new 2-cell, and the original two 2-cells are modified, so that they do not include the intersection area.

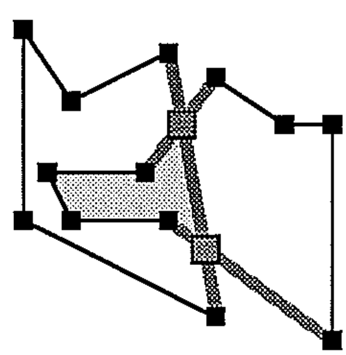

Figure 2: New cells introduced when two 2-cells intersect.

Single Geometry Layer

In contrast to maintaining several thematic layers organizing only a subset of spatial objects, it is recommendable for topological data structures to integrate all spatial data in a single geometry layer*. Spatial analysis requires access to topological relations between different spatial objects. The integration of all spatial data into a single layer of such a topological data structure provides for quick retrieval of all these relations. The integration is equivalent to the preprocessing of all overlays or the precomputing of all intersections [Frank 1987a]. Therefore, fast interactive queries are possible even if intersections of objects or overlays are asked for.

Reason for Performance Problems

The reason for the performance problems of single topological data structures was shown to be the large number of cells contained in large spatial object. This paragraph explains, why large objects contain large numbers of cells. Two factors determine the number of cells describing a spatial object: (1) the degree of spatial resolution chosen for the structure, and (2) the number of other spatial objects that are topologically related to the object.

Topological data structures model all objects in a single spatial resolution, i.e., the geometry represented approximates the real world geometry with equal status for all objects. Objects are defined by component cells and their geometry; each cell approximating a section of the real world geometry. The geometry of a cell is described by a shape type and shape parameters. The number of shape types is typically very limited. For example, only 'straight line segment' and 'arc' are available shape types for 1-cells. Figure 3 shows such an approximation of the real world geometry with straight line segments and arcs as cells.

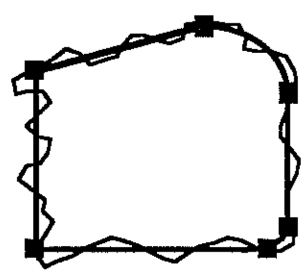

Figure 3: Approximation of the real world geometry with cells.

This way of defining the geometry of objects makes it obvious that an increase in spatial resolution increases the number of cells defining an object. Similarly, a relatively large object, on the average, contains more cells than small objects represented in the same resolution.

The number of cells per object is further determined by other objects that are topologically related to the object. As described before, cells are not allowed to intersect. Instead, the cells must be split into several cells. If objects are topologically related, their component cells as used in the definition of the geometry would intersect. The definition cells can thus not directly be inserted into the structure, but have to be split first. On the average, relatively large objects are topologically related to a large number of other, mostly smaller objects. For example, a state is originally defined by a single 2-cell. It is

* This does not contradict the idea of multiple layers in a multiple representation, as the topological relation of **all** spatial objects are organized in a single structure.

topologically related to all of its counties, all of their communities and all of their parcels. In the topological data structure, the originally single 2-cell is therefore split into a large number of 2-cells that correspond to the parcels contained in the state*.

HIERARCHIES OVER TOPOLOGICAL DATA STRUCTURES

After describing the problems of single topology-based representation, in this chapter a hierarchical structure is proposed to overcome these problems.

Higher Level Cells

The concept of higher level cells is a possible solution to build a multiple representation. Higher level cells are a 'higher level view' of the cells known from single topological representations. Like in other hierarchical structures such as strip trees and quadtrees, the higher level cells are coarser and exhibit less detail. The multiple topological representation proposed consists of several levels of single topological data structures. The lowest level or so called 'base level' is identical to a single topological representation organizing all spatial objects. The cells of this representation shall be called 'base cells.' 'Higher level cells' are cells of a topological data structures on a higher level. Adjacent levels are interconnected over hierarchical relations between cells: The relation maps one cell on the higher level to several cells of the same dimension on the level underneath. A higher level cell can thus be interpreted as an aggregation on lower level cells. Figure 4 shows higher level cells of dimension 1 and 2 with the aggregations that form their lower level equivalent. The higher level 1-cell corresponds to an open polygon of lower level 1-cells, and the higher level 2-cell is built by an aggregation of adjacent lower level 2-cells. A higher level 3-cell (not shown in figure 4) is totally filled with lower level 3-cells.

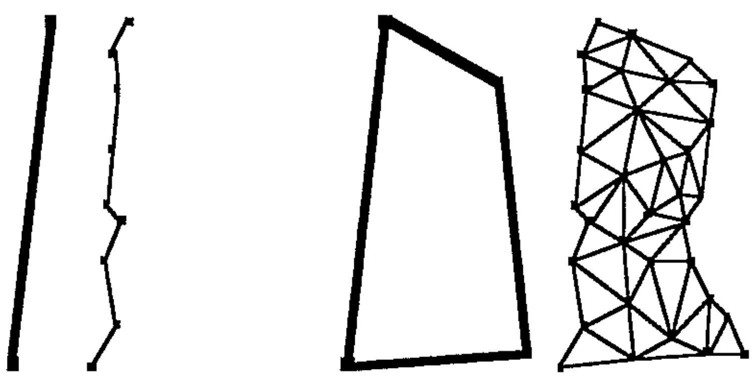

Figure 4 : Aggregates of 1- and 2-cells with their higher level equivalents.

Metric Interpretation

Unlike the cells of the base level, higher level cells do not have a 'natural' metric interpretation. They do not have an actual shape like the base cells that are used to approximate a real world geometry. Because of their relations to lower level cells, however, they can be assigned an imaginary shape called 'composed.' Two metric interpretations of a 'composed' cell are possible: (1) The cell can be interpreted as the simplest convex hull of the metric interpretations of the cell's boundary, i.e., as straight line segment between the end points for 1-cells and as a surface between the bounding edges interpreted as straight line segments for 2-cells. (2) The second interpretation

* If parcels are allowed to lie in more than one community, county or state, the component 2-cells are even finer than those corresponding to parcels.

maps the cell one level down to an aggregation of cells and leaves the decision open how to interpret the cells of this aggregation. In the special case, where a cell is mapped down to the base level, the metric interpretation with the best spatial resolution is found.

Operations to Interconnect Levels

The cells on the different levels are connected by hierarchical relations. As common in the field of abstract algebras [Guttag 1977] these relations will be expressed in the following by a set of operations:

Consists_of is an operation that returns the aggregation of cells of the level below that are equivalent of the argument cell. This operation accepts any cell as argument. If consists_of is applied to a cell of the base level, the argument cell is returned as it cannot be shown in greater detail. If the argument cell is a node, the resulting aggregate will contain a single node of the level below.

The *contained_in* operation expresses the same relation between levels but connects the cells upwards. The cell returned is defined on the the level above and contains the argument cell. It will not always be possible to return a cell of the level above, as is obvious for top level cells and because cells on other levels do not always have a higher level equivalent. In these cases, the argument cell shall be returned, as no equivalent cell of coarser resolution is available.

ShapeSimplification is an operation normally applied to 1- and 2-cells. It returns a measure for the simplification of shape introduced using the metric convex hull interpretation of the cell* instead of mapping the cell down through all levels and considering the equivalent aggregate of the cell on the base level with all its metric information. One possible way to measure the the simplification of shape is to approximate the aggregate of cells on the base level by a box of the appropriate dimension and compare it with the convex hull interpretation. In a two dimensional embedding space this box is a strip that is, perhaps, familiar from the strip tree concept of [Ballard 1981]. The measure of simplification can be expressed by the ratio of the length of the straight-line-interpretation and the strip area. In a three dimensional embedding space, the aggregations of 1-cells and of 2 cells are both approximated by boxes. For 1-cells the measure of simplification is the ratio of box volume and the length of the straight line interpretation. For 2-cells the ratio of box volume and area of the surface interpretation is considered. Figure 5 shows how a strip approximates an aggregation of 1-cells. Figure 6 demonstrates how the convex hull interpretation of a higher level cell is compared to the aggregation on the base level.

* A metric interpretation for 2-cells is only of interest in three-dimensional embedding space. The method of the cell complexes is the only topological representation referenced in this paper that supports the third dimension. As 2-cells in cell complices are always triangles, the metric interpretation is a simple plane segment.

* The convex hull interpretation is a straight line segment for a 1-cell and a surface for a 2-cell.

Figure 5: Strip approximating an aggregation of edges.

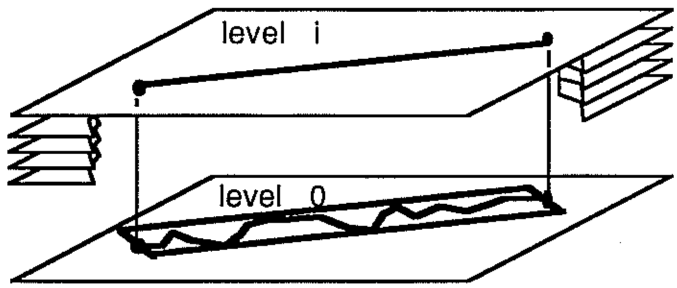

Figure 6: Convex hull interpretation of a 1-cell and corresponding aggregation of 1-cells on the base level

Specification of Operations

A system dealing with the cells of a multiple topological representation can be entirely described by a formal algebraic specification of its operations. The operations can be divided into those dealing with cells within a level and those interconnecting levels. The first group of operations is the same as in a 'classical' topological data structure. Frank and Kuhn have specified these operations formally [Frank 1986]. In this paragraph, an algebraic specification of the operations to interconnect levels is presented. The axioms describing their semantics have not been included, as the semantics have been sufficiently described in the last paragraph.

The following terminology will be used in the specification: *i:d*-cell refers to a general higher level cell and is read as 'level i, d-dimensional cell'. *i* is an integer indicating on which level of the structure the cell is defined. The base level is equivalent to level 0. A 2:3-cell, for example, indicates that the cell is a 3-cell (i.e., a volume) defined on the third level of the hierarchy, i.e., after aggregating twice.

The following operations are necessary to interconnect levels in multiple topological representations. The formulation is dimension independent, i.e., any dimension is allowed for the dimension parameter d.

consists_of	:	i:d-cell	->	aggregation of (i-1):d-cells
contained_in	:	i:d-cell	->	(i+1):d-cell
shapeSimplification	:	i:d-cell	->	simplificationMeasure

Multiple Representation of Objects

A description of one possible organization of objects follows. The geometry of spatial objects in this multiple topological representation is defined as aggregation of cells on a certain higher level. Implicitly, the objects are represented on all levels below this definition level making use of the hierarchical relations between cells.

On the base level, spatial objects are available in maximal resolution and detail. On the first level, objects can be represented without the detail introduced through the initial metric approximation of the real world geometry with cells. Because of the very limited number of shapes available, sections of the object geometry had to be split into several sections in order to approximate their geometry with the cells available. The first level simulates arbitrarily shaped cells by undoing all these splits introduced for the metric approximation. Of course, all the splits introduced by intersections of objects have to be maintained. This relationship between the base level and the first level determines which base level cells have to be aggregated to form first level equivalents. All objects are still represented on this first level.

For determining the second and higher levels, an order relation can be found that expresses the importance of object classes. Small spatial objects are usually relatively unimportant compared to large objects. The order relation can be used to determine the level on which an object shall be defined. The object is represented on this level and all levels underneath, but not on the levels above. Small objects and details thus do not exist on high levels. The number of objects decreases with the height in the structure. Because fewer objects produce fewer intersections, the component cells of the objects are split less and less on higher levels. The order relation over the objects obviously determines which cells have to be aggregated to form higher level equivalents.

Multiple Resolution

The operations discussed allows views of objects in different spatial resolutions. The key to this possibility are the two metric interpretations available for 'composed' cells. Using the aggregation-interpretation, every component cell of the object can - starting on the highest level - be mapped down to a level where the spatial resolution is appropriate. On this level, the convex-hull-interpretation is applied for displaying the cell. For example, the state of Maine shall be displayed on the screen. The system finds the edges of the state boundary on the highest level and maps each of them down through the levels using the 'consists_of' operation until 'shapeSimplification' indicates that sufficient resolution for the display is reached. This will typically happen on different levels for different edges of the boundary.

Performance

The performance problems observed with single topological representations should not occur in the proposed multiple representation. Large objects can now be handled on higher levels, where they consist of fewer cells and can thus be handled efficiently. The level chosen depends on the level of detail and the spatial resolution required.

CONCLUSIONS

The proposed multiple representation supports multiple spatial resolution and provides higher levels where large objects can be described with a small number of cells. The problems observed in single topological representations thus cannot occur in this multiple representation. This opens the possibility for the implementation of multipurpose GIS where several user groups share their data. This allows several user groups to acquire and update their data in a common effort and thus avoid considerable costs for duplication of work.

The structure proposed in this paper is a special case of a multiple representation, the topic of Research Initiative 3 of the National Center for Geographic Information and Analysis [Abler 1987]. This research helped in gaining a better understanding of the problem and will be continued.

The continuation will address the organization of objects in the proposed structure in more detail. Procedures to build, modify, and update the structure without introducing inconsistency have to be found. Further, query processing and optimization for this structure must be dealt with.

ACKNOWLEDGEMENT

Max Egenhofer's comments on several versions of this paper are highly appreciated. Thanks to Kate Beard and Renato Barrera for their stimulating discussions on the topic.

REFERENCES

[Abler 1987] R. Abler. The National Science Foundation National Center for Geographic Information and Analysis. International Journal of Geographic Information Systems, Vol. 1, No. 4, 1987.

[Ballard 1981] D. H. Ballard. Strip-Trees: A Hierarchical Representation for Curves. Communications of the ACM, Vol. 24, No. 5, 1981.

[Boudriault 1988] G. Boudriault. Topology in the TIGER File. In: N. R. Chrisman, editor, Auto Carto 8 Proceedings, Baltimore (MD), 1987.

[Bruegger 1989] B. P. Bruegger, Renato Barrera, A. U. Frank, K. Beard, and M. Ehlers. Research Topics on Multiple Representations: Position Paper for NCGIA Initiative 3 . NCGIA, Initiative 3, forthcoming, 1989.

[Corbett 1979] J. P. Corbett. Topological Principles of Cartography. Technical Report 48. Bureau of the Census, Department of Commerce, 1979.

[Frank 1983] A. U. Frank. Data Structures for Land Information Systems--Semantical, Topological, and Spatial Relations in Data of Geo-Sciences (in German). Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland, 1983.

[Frank 1986] A. U. Frank and W. Kuhn. Cell Graphs: A Provable Correct Method for the Storage of Geometry. In: Proceedings of the 2nd International Symposium on Spatial Data Handling, Seattle (WA), 1986.

[Frank 1987a] A. U. Frank. Overlay Processing in Spatial Information Systems. In: N. R. Chrisman, editor, Auto Carto 8 Proceedings, Baltimore (MD), 1987.

[Frank 1987b] A. U. Frank. Towards a Spatial Theory. IGIS Conference, Crystal City, November 1987.

[Guttag 1977] J. Guttag. Abstract Data Types and the Development of Data Structures. Communications of the ACM, June 1977.

[Herring 1988] J. R. Herring. TIGRIS: Topologically Integrated Geographic Information System. In: N. R. Chrisman, editor, Auto Carto 8 Proceedings, Baltimore (MD), 1987.

[Jackson 1987] J. Jackson. Algorithms for Triangular Irregular Networks based on Simplicial Complex Theory. In: Technical Papers of the 1989 ASPRS-ACSM Annual Convention, Baltimore (MD), 1989.

[Jones 1986] C. B. Jones and I. M. Abraham. Design Considerations for a Scale-Independent Cartographic Database. In: Proceedings of the 2nd International Symposium on Spatial Data Handling, Seattle (WA), 1986.

[Kinnear 1988] C. Kinnear. The TIGER Structure. In: N. R. Chrisman, editor, Auto Carto 8 Proceedings, Baltimore (MD), 1987.

[Meixler 1985] D. Meixler and A. Saalfeld. Storing, Retrieving and Maintaining Information on Geographic Structures: A Geographic Tabulation Unit Base (GTUB) Approach. In: Auto Carto 7 Proceedings, Washington (D.C.), 1985.

[Pullar 1988] D. V. Pullar. An Overall Approach to Data Modeling for Applications in Geographical Information Systems. Master Thesis, University of Maine, Surveying Engineering, Orono (ME), 1988.

[Samet 1984] H. Samet. The Quadtree and Related Hierarchical Data Structures. ACM Computing Surveys, Vol. 16, No. 2, June 1984.

[White 1985] M. White. A Survey of the Mathematics of Maps. Auto Carto 7 Proceedings, Washington (D.C.), 1985.