WEB PRESENTATION OF SPATIAL DATA

Kapil Aggarwal[a]*, Rekha Govil[b]

[a] Computer Science Faculty, Dept of Computer Science, Banasthali Vidyapith, Rajasthan, India – (kapil594@rediffmail.com)
[b]Prof. & Head, Dept of Computer Science, Banasthali Vidyapith, Rajasthan, India

KEY WORDS: Spatial object, Visual significance, Applet, Spatial data set generalization, Object generalization, Minimum bounding rectangle, Spatial indexing.

ABSTRACT:

In spatial applications, it is often necessary to use spatial objects with varying degree of detail. As spatial objects are typically displayed in a way that allows human users not to discern unnecessary details, it would suffice to draw an abstraction of the spatial data preserving its characteristics. Generalization, the process to derive such a less detailed representation, may lead to a remarkable reduction of the computational overhead involved with displaying complex spatial objects. It can also benefit data transfer from one site to another. With a growing number of interactive spatial applications on the web, spatial data generalization becomes increasingly important. In this paper, we investigate how database systems can support generalization of spatial data. We introduce the concept of visual significance of spatial objects for interactive spatial applications, and suggest efficient ways to produce a good quality generalized map for web-based spatial applications.

1 INTRODUCTION

Spatial database systems (SDBS) and Geographic Information Systems as their most important application aim at storing, retrieving, manipulating, querying, and analyzing geometric data. During the last two decades, spatial database systems have become a vital part of Geographical Information Systems (GIS) for storing and accessing spatial data. Spatial Databases is the first unified, in-depth treatment of special techniques for dealing with spatial data, particularly in the field of geographic information systems (GIS). Spatial database systems offer the underlying database technology for geographic information systems and other applications. A spatial database is a specialized type of relational database which is optimized to store and query geographic data, including points, lines and polygons. While typical databases can understand various numeric and character types of data, additional functionality needs to be added for databases to process spatial data types such as shapes. Their role is pronounced by the steadily increasing amount of spatial data maintained in such systems. While the data stored in a database represents the finest level of details available, for a given application it is often desirable to use a level of details suitable for the application. The derivation of an abstract representation is known as *generalization*. Generalization derives from a source dataset a target dataset at a reduced scale whose contents and complexity have been reduced in such a way that the structural characteristics of the source data are maintained for a given application. By removing excessive and non-relevant details, it is possible to derive a representation of the source data that is much more suitable for the given application scenario. Generalization can reduce the data volume considerably since a target data set typically consists of fewer and simpler data objects than the source data. This performance aspect of generalization becomes increasingly important with interactive spatial applications using spatial vector data on the Internet. Over past two decades, tremendous research efforts have been spent in GIS community on the knowledge acquisition for automatic generalization based on the rule-based systems.

A general architecture for Internet-based spatial applications uses a spatial database to store both spatial and non-spatial data. The database can be browsed by means of a Web-browser capable of running Java applets. Generating requests and displaying the result data is handled by the applet that runs on the user's machine. On the server site, a database web server translates the user's request into queries against the spatial database to fetch qualifying objects. These objects are then encoded according to some spatial data transfer protocol and transferred to the applet over the network. After decoding the data on the client side, they are drawn on the screen. Generalization can improve the performance on both the database side & the application side. Obviously, once the objects are retrieved from the database, they can be simplified on the server site by removing parts of an individual spatial object or finding a suitable abstraction that preserves its characteristics. This type of simplification is called as *object generalization*. In this paper, more focus is given on a specific type of generalization, called as *data set generalization*. It deals with removing objects from data sets and addresses the issue of spatially significant objects. It approaches the problem of generalization from a database perspective, treating generalization as an integral part of database query processing. A database query generated from a user's request is modified taking into account of how these data are to be used. If some objects cannot be perceived when being displayed in the web browser, data set generalization tries not to retrieve them from the database at all. However, generalization can degrade the quality of map if the characteristics of the original map are not preserved during generalization. In addition, the data set generalization step may also increase the overall processing costs. Thus, it is crucial to find an efficient data generalization method that can significantly reduce the amount of data, and at the same maintain the key characteristics of the original data set, with as little as possible overhead. Spatial objects are typically large. Therefore, it is a waste of resources if some objects are fetched from the database but are found later as insignificant to the application. Based on this observation, one can introduce a concept of *visual significance*, and a method to determine the visual significance of spatial objects by examining their index entries rather than the objects themselves. In other words, the full geometry of a qualified object is only retrieved from the database when it is likely to be used by the application. Checking whether an object is significant at the index level can be done efficiently because index entries are much smaller in size and simpler in structure in comparison to the objects they represent.

2 BACKGROUND

2.1 Generalization of Spatial Objects

Database derivation is a useful way in building a new database in GIS. It is one of the application fields, which need generalization facilities. Generalization in GIS transforms data into an adequate one to be represented at a smaller scale. It means that generalization derives new data through transformation of the spatial and non-spatial properties of a feature. A feature contains four elements including geometry, non-spatial properties, topological relations and non-topological relations. These elements are changed through generalization. First, a generalization operator changes the geometry of a feature where it generates scale-dependent data. This generalization has been called map generalization or cartographic generalization, since it concerns only geometric changes. Secondly when geometry is changed, topological relations, non-topological relations or both can be changed as well. The generalization must preserve consistency of topological relationships and it may create, delete or derive features and non-topological relationships. When a non-spatial property is changed, it causes a set of changes of geometry; topological relationships or non-topological relationships can be changed. These generalizations are called model-oriented generalization, since it concerns changes of data model in GIS databases.

Generalization is difficult because there is no unique solution, but numerous constraints have to be taken into account during its process. The constraints that hold for spatial data that is to be displayed to humans are like *Metric constraints*, *Topological constraints*, *Semantic constraints etc.* Generalization process consists of some basic operations like Selection, Simplification and Tokenization. It might be necessary to displace some objects to preserve their topological relationship. Data set generalization is the first step of generalization. The operations for simplification, tokenization and amalgamation are applied to the objects selected from the database.

2.2 Visual Significance

The purpose of generalization is to identify for a given data set *significant objects* and *significant parts of the objects*. We distinguish two types of "significance" here. The first one is determined by the semantics of the request. For example, users searching for crop data in a particular region typically have no interest in data irrelevant for their request. Therefore, crop data is considered as significant. For this type of significance, it is possible for the user to *explicitly* specify as attribute constraints, which can be readily used by the underlying database system for object retrieval. This is in contrast to the second type of significance, which deals with *implicit* constraints inferred from the settings of the application and the request. An example of the second type of significance is that for a given application request it is possible to infer constraints that objects smaller than a certain size should not be displayed because they are too small to be discernible. The conditions to describe the second type of significance depend on the intention of the application and the success of generalization highly depends on the kind and quality of the implicit conditions. In the past, it has been argued that human involvement is necessary to find such conditions [6]. Interactive spatial applications, which do not display all the data items meeting the explicit constraints, have to derive automatically implicit conditions allowing the identification of “visually significant” objects. Intuitively, we say that a spatial object is visually significant if it is in one of the following cases: (1) *Large objects*: Objects larger than certain size are important and therefore, they should be selected. Here, size refers not only to area, but also to other criteria such as extension. For example, long major roads with a comparatively small area should be considered as large objects in terms of visual significance. (2) *Objects in sparse region*: Small objects in a sparsely populated region can also be significant. A small town in the desert of central India can be more significant than a suburb in a metropolitan area, even when the latter happens to be considerably larger. (3) *Representative objects*: Within a large set of qualifying objects, it might be necessary to select a subset of objects as representatives for the complete set. For example, while it may not be possible to display all the land parcels in an area, some land parcels need to be displayed to indicate the land use. Note that in order to select representative objects, the objects on the boundary of a cluster of objects might be more important than the object inside. The first criterion in our visual significance definition above is a metric constraint, and the last two criteria attempt to model Gestalt constraints in order to give a better overall feeling about the map.

3 METHODS

When there are too many objects satisfying the explicit query condition, some techniques need to be developed to filter out the unnecessary objects, while allowing the visually significant ones to pass through. These techniques will make use of implicit conditions derived from display parameters. Broadly speaking, four possible techniques can be used for generalization:

(1) *Random selection*: Objects are chosen at random from the source data for display. This solution is discounted because there is no guarantee that the target data set “look” similar to the source data, and there is no consideration about object sizes when removing spatial objects.

(2) *Multi-scale database*: Develop a database that store objects with different scale and generalization levels. Instead of developing the methods to transform to any level of scale/generalization, the different levels are pre-coded in the database. This overcomes the speed problem for the generalization process and removes the need to code the transformations to work on any data and with any input. However, it suffers from two major 4 problems. First, this leads to a huge storage overhead for materializing spatial objects at potentially very large number of scales. This problem can be prohibitive when the spatial data set is very large. Second, this also poses maintenance problems such as maintaining consistency among different representations of the same spatial object when some objects are modified.

(3) *On the fly calculation outside of the spatial database*: All objects are fetched into memory and examined on the fly to determine what should be displayed. Although its accuracy and quality would be high, this approach is problematic because of the excessive time required to fetch and examine each object.

(4) *On the fly calculation inside of the spatial database*: If visually insignificant objects can be identified inside the spatial database, they will not be fetched and processed outside the spatial database. Inside the spatial database there are much information about object approximation and spatial indexing, which can be used to identify significance of spatial objects without retrieving and processing the full geometry of a large portion of objects.

3.1 Object Approximation and Spatial Indexing

Let us briefly introduce the concept of object approximation and spatial indexing used for web presentation of spatial data.

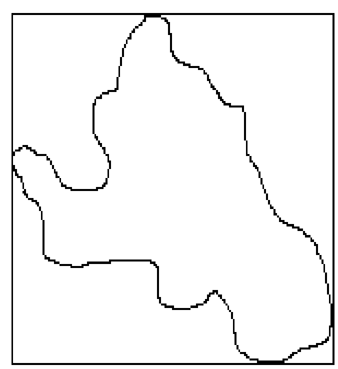

Figure 1: Minimum bounding rectangle

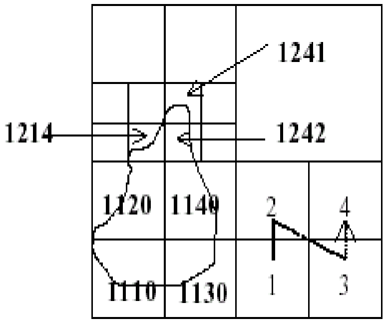

Figure 2: Polygon approximation using z-values

A polygon can be arbitrarily complex. It is a common practice to approximate a spatial object using its *minimum-bounding rectangle* (MBR), which is defined as the smallest rectangle that completely encloses the object, see Figure 1. In such a way, an object with any number of points can be approximated using a rectangle that can be represented by only two points (i.e., the lower-left and upper-right corners). The cost of processing spatial data can be reduced significantly by processing their approximations first. For example, two polygons cannot overlap if their MBRs do not overlap; thus, a simple operation on MBRs may eliminate the need to fetch and process full geometry for some polygons, reducing both I/O and CPU cost.

An object often needs to be approximated at a much finer level than its MBR. For example, the MBR of a long road crossing the data space is obviously an inappropriate approximation. The *z-ordering* technique is based on space filling curves. In Figure 2, the polygon has been approximated by 7 polygons, where the resolution is 4. By transforming a two-dimensional object into a set of one-dimensional points (i.e., the z-values), spatial objects can be represented as numbers and therefore can be maintained, for example, by the B+-tree [1].

3.2 Algorithms used in data set generalization

Three algorithms are chosen to do data set generalization at the database level: (1) *A check for size based on MBR*. The MBR provides a simple way to estimate the size of an object. This estimation can then be used as a measure of the significance of the object. A comparison of an object's MBR area with a threshold provides a way to filter a set of objects based on size, where the threshold can be derived from the spatial extent of all objects to be selected against the size of the display area. This algorithm assigns visual significance based on the size of the object. (2) *A check for size based on z-value information*. Using the z-value that describes an object in a spatial index can be used as an estimate of the relative size of the object. This algorithm fetches the z-values that describe the location of an object and examine their length as well as the total number of z-values for the object. This algorithm assigns visual significance based on the physical and spatial size of the object. (3) *A check of isolation based on z-value information*. By dividing the requested region up into fixed size areas, each can be examined to see how many objects exist in that region. If there are below a certain number, all objects in the region are deemed isolated and hence visually significant. For regions with a total number of objects above a certain size, uniform filtering can be performed to reduce the number of objects while maintaining a good impression of the original map. Z-values are used to fetch objects in a given area. From this the amount of objects that exist in this area of space can be determined. This allows a measurement of the isolation of the object to be made. This algorithm assigns visual significance based on the isolation of an object. If a region has below a certain amount of objects all the objects in that region are included.

4 EVALUATION CRITERIA

The above three methods are further to be evaluated and compared on the basis of (a) visual quality of generalized maps and (b) time required in producing them. Out of these three methods, the best method is picked, which requires less time than the method of no generalization due to a smaller amount of target data it produces. It is also be able to produce the best quality map, which satisfies all the criteria we used to define visual significance.

5 CONCUSIONS

Currently available web-based spatial applications face major performance problems when it comes to displaying vector data as extracted from a spatial database system. The performance of these systems is well below their potential and there is a pressing need to develop efficient techniques to achieve a performance close to conventional systems. Among many possibilities to improve the performance of web-based spatial applications, the most obvious one is to avoid processing data not used by the applications. It is possible to provide database support for generalization by identifying irrelevant data already at a very early stage. We have demonstrated that some limited form of generalization can be performed by means of information stored in a typical spatial index. This information allows us to produce a simplified map with satisfactory visual quality with no increase of data retrieval time.

6 REFERENCES

[1] D. J. Abel. SIRO-DBMS: A Database Toolkit for Geographical Information Systems. *Int. J. Geographical Information Systems*, 4(3): 443-464, 1989.

[2] V. Gaede and O. Günther. Multidimensional Access Methods. *ACM Computing Surveys*, 30(2):170-231, 1998.

[3] R. B. McMaster and K. S. Shea. Generalization in Cartography. Association of American Geographers, Washington, D. C., 1992.

[4] Map Generalization in the Context of Digital Systems. Guest Editorial to Special Issue on Automated Map Generalization, Cartography and Geographic Information Systems, 22(4): 3-10. Weibel, R. (1996):

[5] R. Weibel. Generalization of Spatial Data. In *CISM Advanced School on Algorithmic Foundations of Geographical Information Systems*, pages 346-367, 1996.

[6] R. Weibel. A Topology of Constraints to Line Simplification. In *Proceedings of the 7th International Symposium on Spatial Data Handling (SDH96)*, pages 9A.1-9A.14, 1996.

[7] Bertolotto, M. and M.J. Egenhofer (1999). Progressive Vector Transmission. *ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO.

[8] Buttenfield, B.P. (1999). Sharing Vector Geospatial Date on the Internet. *Proceedings 19th International Cartographic Conference*, Ottawa, Canada.

[9] Agrawal R., Imielinski T., Swami A.: *"Database Mining: A Performance Perspective"*, IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, 1993, pp. 914-925.

[10] Egenhofer M. J.: *"Reasoning about Binary Topological Relations"*, Proc. 2nd Int. Symp. on Large Spatial Databases, Zurich, Switzerland, 1991, pp. 143-160.

[11] Ester M., Kriegel H. -P., Sander J.: *"Spatial Data Mining: A Database Approach"*, Proc. 5th Int. Symp. on Large Spatial Databases, Berlin, Germany, 1997, pp. 47-66.

[12] Conceptual framework for automated generalization and its application to geologic and soil maps. Conference Paper, ICA Conference, 2005, La Coruna, Spain.

[13] J.P.Lagrange, A.Ruas, *Geographic Information Modelling: GIS and Generalisation*, SDH, 1994, pp1100-1103.

[14] John.G.Stell, and Michael.F.Worboys, *Generalizing Graphs Using Amalgamation and Selection*, SSD, 1999.

[15] R.B.McMaster, K.S.Shea, *Generalization in Digital Cartography*, The association of american geographers, 1992, p3.

[16] Esri, Automation of Map Generalization, 1996, ESRI.