GLOBAL INDEXING OF 3D VECTOR GEOGRAPHIC FEATURES

J. Kolar

Centre for 3DGI, Aalborg University, Jernes Niels Vej 14, 9220, Denmark, kolda@3dgi.dk

KEY WORDS: Spatial Information Sciences, Global, Multiresolution, Three-dimensional, Databases, Geography, Geometry.

ABSTRACT:

Geographic data management is essential for development of geo-related applications and systems. Fast access to a subset of data related to an area of interest with a sufficient precision is a common need for geographic applications. The mechanism for organizing data in the database determines how geographic features can be represented, whether or not the underlying database can be distributed or how the database can be updated. This article introduces a uniform global indexing method for 3D vector geographic data. The index facilitates several properties important for geographic systems, such as fast access to features with a similar geographic location, a multi-resolution representation, a global uniform coordinate system and the avoidance of using cartographic projections. In order to avoid truly 3 dimensional spatial indexing the solution is based on a tessellation of the space using two spherical coordinates. This decreases the computational complexity and increases performance. The article presents description of main concepts of the indexing method. Introduced method has been implemented and used with real terrain data. Also example applications are considered and practical connection with an RDBMS is proposed.

1. INTRODUCTION

1.1 Global Grids

Indexing spatial data globally provides a unique approach for organizing geographic or any geo-related data. Various works on global grids has been elaborated in order to facilitate management of vector data, raster data, data quality and resolution of data. The main motivation for the presented indexing technique is development of referential model of Earth suitable for visual navigation. Taking this effort from the scratch an application of global grids as an indexing approach is a very essential issue of such development. However, not all spatial data management aspects of global geographic applications have been considered and even more left to be solved. Global grids, as division schemes, are naturally suitable for indexing spatial data.

The major indexing property of global grids is division of the space into smaller areas. The task of an indexing application then is to access data from these areas fast, even though there are possibly huge amounts of data for the rest of the spatial domain covered by the grid.

Another important issue is ability to edit data from these smaller areas independently from the rest of data in the covered domain; at least to certain extent. This means for example to have possibility to edit parts of the model representing Northern Jutland in Denmark independently from the part of the same model covering Sao Paolo in Brazil. Assuming that it is possible to do this using the same technology would have advantageous implications for distributed development and maintenance of global geographic model.

The possibility of the maintenance “in pieces” would also facilitate constructing the model gradually, which is an important requirement for practice. Global grids also influence how the resulting geographic model can be exploited and shared by other geographic applications.

A support for multiple level of detail of the model at various scales can be one such influence global grids can be behind. And it is particularly significant for visual applications. Proposed indexing technique is focused on visualization of global vector data, while most of the works were oriented towards raster data and data projected on a map plane.

1.2 Related Work

Several works elaborating different types of global grids have been published.

In (Dutton, 1989) is introduced an irregular grid that divides sphere using recursive division of faces of octahedral into triangles. The quaternary triangulated mesh (QTM) method divides each face of octahedral into four same triangles whose vertices are projected on sphere afterwards. The process of division can be performed recursively. QTM has been used for various applications such as location code that allows locating any position around the globe together with its precision using only one number (the code). Also applications for map generalization, indexing and terrain representation have been elaborated.

The global grid presented in (Lukatela. 1987) is based on Voronoi diagrams on the sphere. The shape of the division scheme is given by centroids distributed around the origin. The tessellation is based on radial proximity to the centroids. The scheme has been devised as one level global indexing technique for spatial data. The division scheme has been used for terrain representation too.

In (Aasgard, 2002) is elaborated a solution for projecting 2D regular grid on sphere and thus take advantage of quad division in global 3D models. This approach is advantageous mainly for raster data but can be used for indexing of any geographic features around the globe. The disadvantage is singularity at the poles and need for projecting all the data before actual indexing.

There are other applications of quad division, e.g., dividing faces of cube or two sides of plane, which generate global grids. Some of them are used in commercial applications.

2. GEOGRAPHIC INDEX

This section presents an original indexing method for spatial data based on a global grid. The primary focus is on explanation of the concepts behind the tessellation itself, i.e., on construction and geometry of the grid.

2.1 Basic Division Scheme Properties

The main concept of geographic index (*geoindex*) is a tessellation of the sphere into *cells* of similar size. This is achieved using Voronoi diagrams on the sphere. While (Lukatela, 1987) assumes an arbitrary constellation of cells, typically derived from the spatial distribution of the data, in geoindex the constellation of the cells is fixed. If the fixed constellation can be described by reasonable simple algorithm then it can be introduced as a general indexing method that could be applied globally. Another interesting implication would be a possibility to index at different resolutions of the division scheme. This is in general impossible with Voronoi diagrams because the cells cannot be divided recursively. The recursion however, can be replaced by an additional parameter that specifies the resolution of the tessellation.

The grid used by geoindex can be presented as a set of specific instances of Voronoi on the sphere; where the instances are given by a *distribution rule*. The distribution rule is used in order to set a convenient density of the grid that would correspond to the nature of the data or to needs of an application.

Although the division scheme is convenient to depict as a tessellation of the surface of sphere, based on radial proximity the grid actually divides completely the whole 3D space around the origin. This twofold approach to the tessellation is convenient because many geographic data and applications do not need three spatial dimensions. For such applications range from the origin is omitted which results in a two dimensional space given by sphere surface. Still however, the sphere itself is obviously defined in 3D space, which provides a good basis for fully 3D geographic applications.

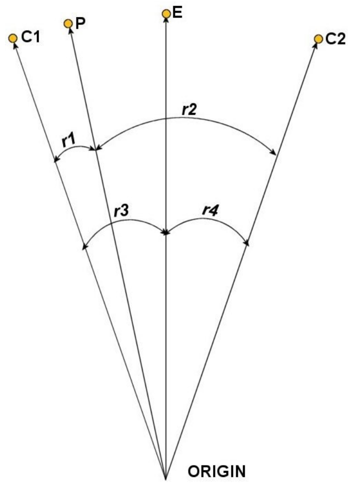

Figure 1. Radial distances

2.2 Cells

The spatial unit in geoindex is a cell. Cell is an essential entity of the method that is directly used in the process of indexing.

In order to define a cell, significant vectors called *centroids* have to be known for each cell. In general, a centroid can be any non-zero vector in 3D. The cell is then given by a set of points with radial distance to a given centroid smaller than to any other centroid as depicted. The situation is depicted in Figure 1 and 2. Cell is identified by spherical latitude and longitude of the centroid. This definition for cell is valid for any point in 3D space.

Cell definition, however, can be interpreted in an analogous manner on the surface of the unit sphere using distance along the surface instead of radial proximity. It is important to note that in the first, more general, definition cells have no metric extents, only radial.

As depicted in Figure 2 and 3 cells have *vertices* and *edges*. Edges are given by vectors from the origin with the same radial distance to two centroids. Vertices are then vectors with the same radial distance to, at least, three centroids.

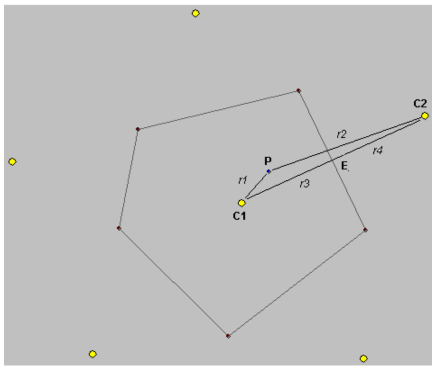

Figure 2. Cell definition

The main task behind the indexing process is to classify a given 3D point to its cell, e.g., to find its *proximate centroid*. In order to identify the proximate centroid it is necessary to know complete set of the centroids from which the closest one can be selected. The very idea in geoindex is that the minimum set of *candidate centroids* can be obtained computationally.

The selection of the candidates is closely bound with the distribution rule for the centroids. The centroids are distributed through points given in spherical coordinates. The points are distributed semi-regularly around the sphere along parallels as follows: Interval between parallels is given by a *division coefficient*. This evenly marks off the specified number of points between poles along the prime meridian as depicted in Figure 4. Parallel at each marked point is then divided using as many points so that the interval along the parallel fits best to the interval along the meridian. Centroids then are defined as set of vectors from the origin to each point marked on the sphere.

Obtaining candidate centroids is based on the following predicate; there are four candidate centroids for an arbitrary 3D point that does not define a centroid. Given such 3D point, its candidate centroids are from the 2 closest parallels---one northern from the point and one to the south. Northern parallel is taken when the point occurs exactly on one of the parallels.

From each parallel are taken two centroids with values of spherical longitude closest to the point. From the four candidates the proximate centroid is found. If two or more candidates have the same distance to the point the most north east centroid is selected. Note that next to the poles only three candidates are available.

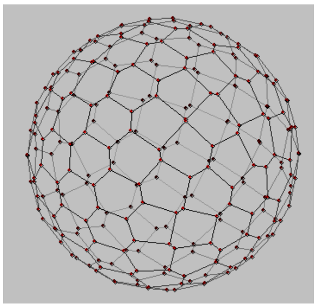

Fig. 3. Geoindex tesselation.

An example of full tessellation of geoindex projected on the sphere is shown in Figure 3. The tessellation generates a semi-regular global grid. The geometry of the cells is not unified but it is possible to claim that cells tend to have a shape of hexagons.

2.3 Levels

In order to use Voronoi based grid for data with various resolutions geoindex introduces concept of *levels*. In applications of geoindex an arbitrary number of levels can be used. Each level tessellates the space into cells as introduced in the previous section. Each level has different division coefficient (see Section 2.2). The levels are accessed directly as independent layers rather than hierarchy. In Figure 4 are depicted cells from levels with division coefficients 10 and 50. For sake of clarity only cells along the prime meridian have been rendered.

There is a straightforward way of constructing an acyclic graph when traversing from the coarsest level to a finer level; however this option has not been elaborated. Building a hierarchical structure across the levels is not necessary since the proximate cell can be accessed directly regardless the resolution of the grid at the particular level. Using independent levels also provides flexibility in decision about which levels are needed or convenient to use.

Level	*Number of cells*	*~width [m]*
2	6	10 000 000
3	12	6 680 000
4	22	5 010 000
5	34	4 010 000
10	128	2 000 000
100	12 732	200 000
1000	1 273 248	20 000
10 000	127 323 974	2 000
100 000	12 732 395 370	200

Table 1. Number of cells in selected levels.

In Table 1 are given number of cells in selected levels. Number specifying the level directly refers to the division coefficient. The coarsest available level is with division coefficient being two. This generates a tessellation of the space through six cells, e.g., a cube projected on the sphere. For this case the width of each cell on the Earth's surface would be approximately 10 000 kilometers.

Use of levels has direct use for dealing with level of detail in underlying applications. Levels with higher division coefficient are accessed when more detailed geographic data from smaller spatial range are necessary, while lower division coefficient is used when larger areas need to be available in their spatial context.

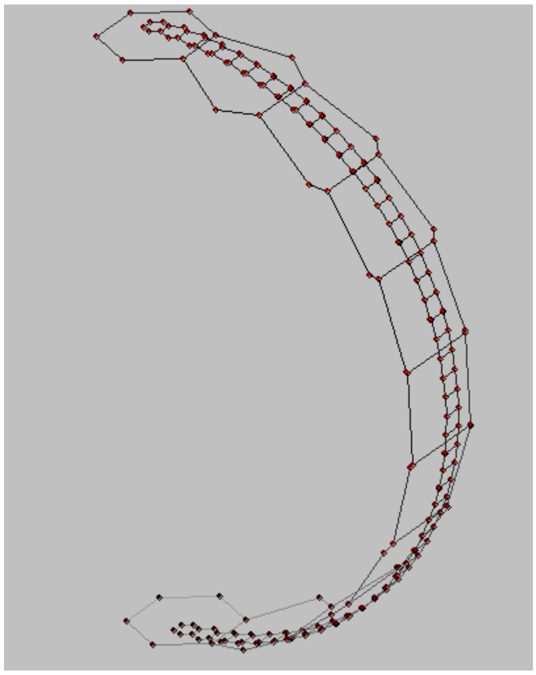

Figure 4. Cells distributed along the prime meridian with division coefficient ten and fifty.

2.4 Applications

In this section is presented how geoindex can be used for management of geographic data in practice. Also reasoning about applications that could exploit the concepts of geoindex is presented.

In order to apply geoindex it is necessary to maintain a search tree and for each record assign an address referencing its physical position in a file or memory. This mechanism, however, is implemented tested and optimized for many years by numerous available database management systems. There are solutions for indexing one-dimensional sets, e.g., B+-tree, which can be used immediately when unique identifiers are assigned to each cell. Similar approach has been proposed in (Oosterom, 1996).

Geoindex has been devised to aid visual applications in 3D. That is one of the main reasons why proximity features of Voronoi diagram have been picked over hierarchical structures as used in QTM or quad division based global grids. Typical query when navigating visually through geographic 3D model can be in a form "give me the nearest data with highest level of detail and distant locations with less detail and from certain distance nothing at all". This closely expresses the concept used by geoindex and any application that can elaborate on this approach can use geoindex.

Regarding geometric data representation that can be indexed using this method; any geometric data structures representing spatial features can be used as long as the data representation supports levels of detail aligned with the levels in geoindex. As an example; the original application and reason for developing geoindex is global terrain representation. The main concepts have been presented in (Kolar 2004).

2.5 Implementation

An implementation of the presented indexing method has been made using Java in a single class. Distribution of centroids given a division coefficient has been coded as a method as well as obtaining proximity cell for a point on a given level. In order to facilitate the application following methods were added: method for encoding and decoding centroid coordinates into unique identifiers; method for computing vertices of the cell; and method for obtaining neighboring cells.

The implementation has been made using double numeric precision (64-bit float number). This limits the division coefficient to be at most equal to 1073741823. Using this division coefficient would mark cells on the Earth surface with extent of approximately two centimeters.

As storage PostgreSQL has been used. PostgreSQL is an open source object-relational DBMS. For creation and management of the indexing data structures has been used B+-tree provided by PostgreSQL. Communication with the database has been made through JDBC.

The data used where GTOPO30 datasets that is freely available as global DTM. For sake of providing notion of performance time measurements from processing 28800000 points (GTOPO tile) are provided. Measurements were made with the prototype implementation and PostgeSQL running on 2.5GHz Xeon machine with 1GB RAM.

Processing the points using geoindex and storing them in PostgreSQL lasts 58 minutes and 28 seconds. Building an index in PostgreSQL using B+-tree and vacuum analyze take 14 minutes and 52 seconds. Note that in the first measurement also other processing than using geoindex is included, e.g. parsing and coordinate transformation. These results are assumed only for better imagination.

3. CONCLUSION

This article overviewed the main concepts of geographic indexing---a new spatial indexing technique using global grids. Global grids provide a uniform approach to "divide and conquer" manner for management of geographic data. Geoindex is a generally applicable indexing method that could be applied globally with a possibility to index with virtually arbitrary spatial resolution.

The used global grid is based on Voronoi tessellation. This is based on a simple algorithm for distributing semi-regular cells around the origin that has been also described. Cells that are used for indexing are defined algorithmically. Hence, in order to use geoindex there is no need to instantiate any cells, e.g., to store any data about them in the memory or on a storage device.

Geoindex can be seen as a spatial extension for indexing techniques used by ordinary RDBMS. The spatial division and mapping to the identifiers can be done by geoindex as a separate process while the actual search tree, e.g., B+-tree, is constructed and maintained inside RDBMS, where our data can be stored. This provides many possibilities for use with existing systems.

Although recursive division of the global grid is impossible, the algorithmic definition and distribution of the cells allows dealing with multiple levels of the grid. Regardless number of levels geoindex performs search for the proximity cell in constant time.

Geoindex can be applied to any 3D spatial data or 2D data on surface of the spheroid. However, it has been designed for spatial data distributed in 3D around certain central point, e.g., spatial data on or near to the surface of celestial bodies, e.g. planets.

Strong feature of geoindex is complete elimination of cartographic projection. When using geoindex together with 3D graphics enabled application there is no need to enforce presentation in map plane. This provides a significant simplification of the process between geographic data acquisition and exploitation/consumption of the data. Also underlying deviations common for all cartographic projections can be avoided consequently.

A disadvantage of presented mechanism is its application with rasters, which would be a complex task. This is due to the variations in the shape of the division scheme provided by geoindex. This "unfortunate" characteristic does not allow a straightforward alignment with strictly given spatial structure of raster data. This fact leaves geoindex suitable mainly for vector spatial data. Another disadvantage is that the scheme cannot be divided recursively, although this is de facto compensated by introducing the concept of levels.

In future work the focus will be given to definitions for more queries that can efficiently exploit the indexing approach. Along with queries the main concern is to elaborate specific applications of geoindex for various data representations of geographic features. Of particular interest is representation of the terrain and representation of linear features e.g. roads or rivers. Also application to rasters would deserve more substantial research.

REFERENCES

Aasgaard, R., 2002. Projecting a regular grid onto a sphere or ellipsoid. In: *Advances in Spatial Data Handling, Richardson*, D. and Oosterom, van P., Springer-Verlag, 339-350.

Dutton, G. 1989. Planetary modelling via hierarchical tesselation. In: *Proceedings of Ninth International Symposium on Automated Cartography (Auto-Carto 9)*, Bethesda, MD, USA, pp. 462-471.

Kolar, J. 2004. Representation of geographic terrain surface using global indexing. In: *Proceedings of Twelfth International Conference on Geoinformatics*. Gavle, SWE.

Lukatela, H., 1987. Hipparchus geopositioning model: an overview. In: *Proceedings of Eighth International Symposium on Automated Cartography (Auto-Carto 8)*, Baltimore, USA.

Oosterom, van P. and Vijlbrief, T., 1996. The spatial location code. In: *The Seventh International Symposium on Spatial Data Handling*. Delft, the Netherlands.