Improving Data Mining of Multi-dimension Objects Using a Hybrid Database and Visualization System

Yan Xia, Anthony Tung Shuen Ho
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798, e-mail: eyxia@ntu.edu.sg

ABSTRACT

The automatic storage and extraction of high-level information characters of three-dimension entity is important to Geoinformation Visualization system. To address this problem, an effort has been taken to develop a three-dimension visualization system integrated with a database when we are focusing on depicting large information spaces. Instead of using serial separated files, we explore the issues in the integration of database with the visualization system; make the database as the kernel of the system to specialize in the storage and management of all types of data (including geographic data). Although database management system does not have the analytical and visualization capabilities of the visualization system, it plays an integral part in our visualization system due to its data management capabilities, and helps to provide the user with a single data model.

Firstly, the paper gives a general overview of the visualization system, and then describes the constitutions and the basic functions of the system. Secondly, the paper focuses on the data structure of the system. Based on the information characteristics of 3D entity in the object-oriented approach, the paper discusses the classification of spatial data, explains the topology relationship among the data, and gives the definition of data structures from macro-structure and microstructure. Thirdly, the paper presents the architectural design of visualization system database, which will store and handle 3D object data. With the integration of database, the end users can specify what to visualize while the system determines how the visualization is to be performed. Therefore, the end users will be able to concentrate more on analyzing data rather than on the processes of gathering and visualizing it.

Keyword: Visualization system, database, data model, efficient database, hybrid database and visualization system

1 INTRODUCTION

Cartography has always been about visualization - in the sense of making aspects of our world visible. The nature of maps and of their use in science and society is now in the midst of remarkable change - change that is stimulated by a combination of new scientific and societal needs for georeferenced information and rapidly evolving technologies that can provide that information in innovative ways [1]. A key issue at the heart of this change is the concept of software visualization.

Software visualization is the use of computer graphics and animation to help illustrate and present computer programs, processes, and algorithms, and can help human being understand how algorithms work and what their codes mean [2]. The Geoinformation Visualization system, named Visual, is such a software visualization system. It provides directly toward the integration of dynamic representations within the sciences for which georeferenced representations are critical (geography, geology, hydrology, and others). The key objectives of the Visual system are development of innovative approaches to and applications of dynamic cartographic visualization and the sharing of perspectives on visualization with the geo-scientific community.

The Visual system provides visualization from creating insightful images to building complete applications for geo-science, it can let Geo-scientists easily gain meaningful insight into the abstract and complex geo-information, and then interpret and understand them. For a high efficient visualization system, the automatic storage and extraction of high-level information characters of three-dimension entity is an important issue. To address this problem, an effort has been taken to develop a three-dimension visualization system integrated with a database when we are focusing on depicting large

Visualization of Temporal and Spatial Data for Civilian and Defense Applications, Glenn O. Allgood, Nickolas L. Faust, Editors, Proceedings of SPIE Vol. 4368 (2001) © 2001 SPIE · 0277-786X/01/$15.00

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms

information spaces, which often contain more informational elements than there are pixels on the screen. We explore the issues in the integration of database with the Visual system instead of using serial separated files; make the database as the kernel of the system to specialize in the storage and management of all types of data. Although database management system does not have the analytical and visualization capabilities of the visualization system, it plays an integral part in our visualization system due to its data management capabilities, and helps to provide the user with a single data model. The paper gives a general overview of the visualization system, and describes the constitutions and the basic functions of the system. The paper then focuses on the data structure of the system. Based on the information characteristics of 3D entity in the object-oriented approach, the paper discusses the classification of spatial data, explains the topology relationship among the data, and gives the definition of data structures from macro-structure and microstructure. Finally, the paper presents the architectural design of visualization system database. With the integration of this database, the end users can specify what to visualize while the Visual system determines how the visualization is to be performed. Therefore, the end users will be able to concentrate more on analyzing data rather than on the processes of data gathering and visualization.

2 VISUALIZATION SYSTEM

The Visual system is an object-oriented Window application written in C++. It is an interactive scientific visualization tool that displays complex structures within three-dimensional geographically referenced data fields [3]. It integrates common database operations and mechanism with the geo-science analysis and visualization capabilities of maps. It is capable of assembling, storing, manipulating, analyzing, and displaying geo-referenced information, i.e. data identified according to their locations.

2.1 Visualization Sub-system

In essence, the Visual system is a collection of methods for mapping and analyzing the spatial properties and potential relationships of objects and events in geosciences. The visual sub-system provides both color shaded-surface display and simple volumetric rendering in either index or true color. Currently supported 3D visualizations are the following:

- Mon-Geological Attribute Surface Model
  For generating and manipulating mon-geological attribute surface, the model routine first computers the scattered data set that describes a desired surface, e.g. a stratum, topographic surface, within the 3D data volume. These data are then rendered as continuously shaded surface. Figure 1 illustrates a sample of a 3D mon-geological attribute surface model view color-coded by elevation to give extra depth and clarity to the true definition of the surface. The Visual contains a wide variety of options that control lighting, viewing, scaling and shading. It is also easy to rotate the 3D surface for optimum viewing from any angle through the graphic user interface.

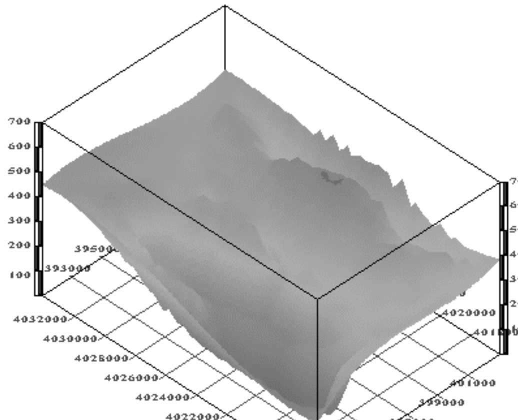

Figure 1 Mon-Geological Attribute Surface Model

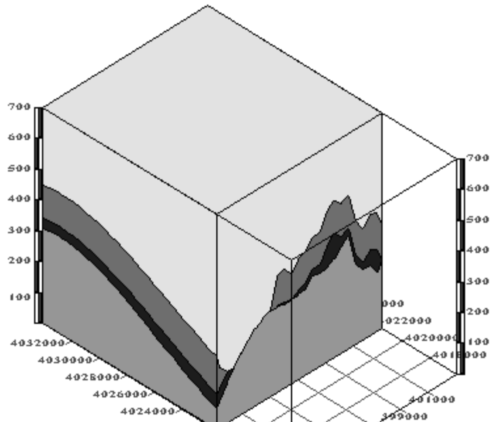

Figure 2 Multi-Geological Volume Model

- Multi-Geological Volume Model
  The model generates, views and plots 3D color-coded perspective views from any 2D or 3D scattered and continuous data set, stratum data, geological structure data, and so on. Objects rendered volumetrically may be

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms

viewed along with shaded surfaces. Additional data sets can be overlaid on shaded surfaces by color-coding the data according to a specified color ramp. 3D visualizations can be viewed in stereo for added depth perspective.

3D data volumes may be sliced in the X, Y, or Z plane using an interactive cutting plane, Figure 2. A cross section of the data volume can be viewed in a 2D window as a 2D contour plot, a vector plot shown in Figure 3, a raster image or a combination of these options superimposed. Also, a portion of the DTM surface in 3-D can be viewed by simply selecting the appropriate area, Figure 4. Map outlines can be used as a background for 2D cross section plots of geographic data. All data are projected according to the coordinates specified by the user for the cross-section window.

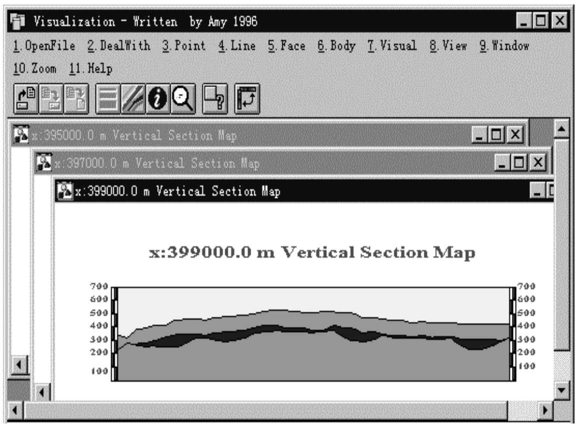

Figure 3 Section

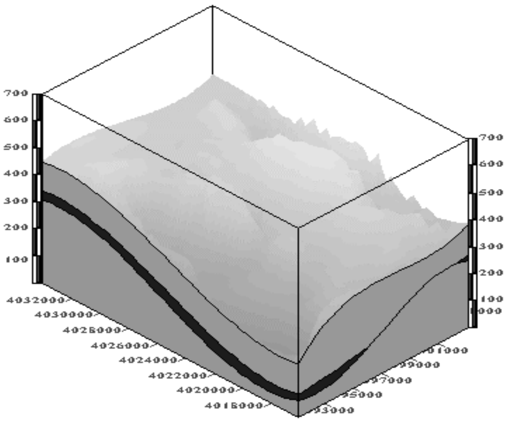

Figure 4 A portion of the DTM surface

Additionally, the graphic user interface provides direct manipulation tools for specifying the eye position, center of view, light sources, and color ramps. Subsetting of data can be performed easily by selecting the data by index or geographic coordinates.

2.2 Database Sub-system

Like data collection, data management and organization is one of the most important components of Visual system. Although geographic reference information can be stored as a series of files within a typical directory structure, the Visual system applies a relational database management system to store, organize, and manage the large amounts of data in order to improve the efficient of the system. Using the relational database model, data is stored as a collection of tables with common fields used to link related records.

Geo-reference Data

The Visual database stores geo-science data (including drill well data, geological data, geographic data, magnetic data, gravity data, etc) about the research area in two separate formats: spatial and thematic. Spatial data describes the location of objects or events. The location can be described by an explicit geographic reference, such as latitude and longitude, or national grid coordinate, x, y coordinates. These geographic references allow user to locate features in the Earth's surface for analysis. Another data format is thematic data. It describes the characteristics of the object or event. It provides attribute information, for example, what name the rock, how depth the drill is, etc. The combination of these two types of data creates a database known as a theme. As such, several themes can be linked and analyzed together by their common geography.

Visual system uses the Gauss-Kruger projective grid coordinate as the spatial data's planimetrical coordinate for analysis or visualization. Figure 1, Figure 2, and Figure 4 show the spatial data's planimeterical coordinates of the objects. So, an automated utility, which converts from original references such as latitude and longitude or polar coordinates to explicit geographic reference – Cartesian coordinates, is a part of the application of the Visual system.

Data and Visualization Interface

The database sub-system provides a data interface to the visualization sub-system. The database sub-system extracts only the relevant data and to export them to a chosen graphic engine in the required format. The data interface is called the Data and Visualization Interface (DVI).

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms

The Data and Visualization Interface addresses a significant problem in the visualization of large data sets. DVI executes as integrated component in the Visual environment, and transfers data between database tables, formats and visualization sub-systems. It has several output ports that are connected to each package, and works within the specifications of each supported format.

Query

Spatial and thematic data are the two basic data types associated with the Visual system database. Spatial data specifies the location of a particular object or event while thematic data provides a detailed description of the object or event. The way that the geo-reference data is obtained falls under one of two possible categories (primary or secondary). Primary data collection involves the design and implementation of a sampling methodology and the subsequent input of the resulting measurements into the Visual system database. Secondary data is generated by the Visual system, such as a surface fitting result, a object's or entity's volume. Once the geo-reference stored within the visual system database, end user can query the information about the object, as shown in **Figure 5**.

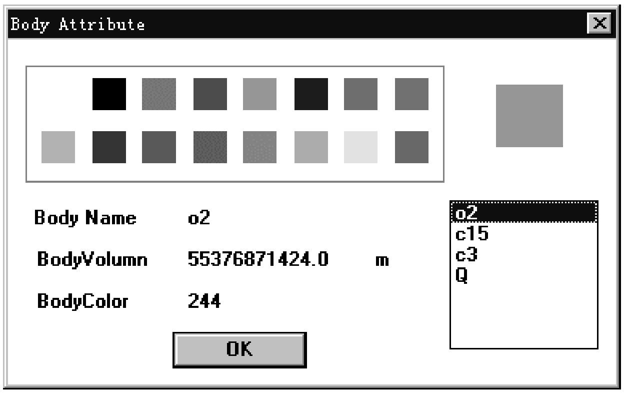

Figure 5 Query result

3 DATA MODEL - 3D VECTOR TOPOLOGY MODEL

Generally, geographic information systems work with fundamentally different types of geographic or spatial models. These models are necessary since computers cannot 'see' the world the same way human do. Therefore, we have to represent real-world objects in a way computer can understand as data structure or data model. Data model is core to the Visual system; it defines both data organization and data handling operations in the system. Different type data in the real world, only organized and integrated according to special data structure, can be stored, accessed, processed and shown easily.

In GIS, there are two important methods for data structure or model definition. The first method is to describe geometric characters of three-dimension geological entity objects via micro-face units or elements. The second method is to represent thematic and geometric characteristics of 3D geological entity object via micro-body units or elements. In the second method, there are also several types of data structure that can be identified: Edge Representation is using the point, line and face to define the position and relationship of the geological entities; Construct Solid Geometry is using the set operation and geometry conversion of simply solid object to define a complex solid; Cube Data Structure is using cubes distributed at space grid nodes to represent 3D geological entity objects; and CAD is using CAD to represent simple 3D geological entity [4]. The Visual system works with a Polyhedron Code method to analysis, organize, store, and visualize 3D georeference entity object [3].

Polyhedron Code belongs to Edge Representation data structure, and is a kind of 3D Vector data structure. According to creating procedure of 3D geological model, it divides the region geological entity to a series of sub-geological entities based on the classification principle of Geology. Each sub-geological entity has thematic and geometric characteristics. The thematic type relationship between sub-geological entity objects is classification hierarchies; the spatial shape of every sub-geology entity is polyhedron. Secondly, Polyhedron Code uses point, line and face three geometric elements to define geometric character (spatial shape) of sub-geological entity. One or more directional enclosed faces define the spatial boundary of each sub-entity. For each surface or face, the boundary is determined by one or more directional enclosed curves, the form and position are defined by the point-coordinates lying on the surface and the boundary, the thematic is defined by the point's thematic. The boundary of the face (surface, out-section, inter-section) is defined by an out-circle. The circle is an enclosed boundary which is composed by a series of sequence and directional edges or arris, the out-circle defines the maximum boundary of face. The form of face is controlled by control-points and can be represented by arcs or circles. The neighboring surfaces of polyhedron object (included sub-body) cross on the edge or arris, the surface of body contains many arcs. A starting node and an ending node controls the direction of these edges and arris. The form of edge or arris is controlled by many control points and defined by many interpolation points. The relationship between body object

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms

and the geometric elements of point, line and face is a multi-hierarchic network in the 3D vector topological model. Polyhedron Code uses the hierarchic structure to manage point, line and face data, and uses the topological relationship of point, line, face to process data.

The Polyhedron Code Data Model is a standards-based, comprehensive, object-specific, geo-spatial 3D model of the metropolitan landscape. It is a 3D vector data structure model, and best suited for dealing with objects or entities whose boundaries are distinct and well defined. It uses coordinates to create point, lines, and face fundamentally three types of geometry element of entity or object. In the Polyhedron Code model, information about points, lines, and faces is encoded and stored as a collection of x, y coordinates. The location of a point feature, such as a bore hole, can be described by a single x,y coordinate, its volume is zero. Linear features, such as faults and drill wells, can be stored as a collection of point coordinates; their volume is also zero. Face features, such as fault faces or research area, can be stored as a closed loop of coordinates. Polyhedron Code uses high geometric precise to represent the position of 3D geological entity in the Euclidean-space, and the topological relationship among data is very clear. The Visual system uses the Polyhedron Code data model to provide end-users with great flexibility in creating visualizations.

4 DATA MODEL STRUCTURE

Polyhedron Code uses high geometric precise to represent the position of 3D geological entity in the Euclidean-space. There are many different concepts for managing data (network, hierarchical, relational, object-oriented), we apply the relational database in Visual system. Using the relational database model, data is stored as a collection of tables with common fields used to link related records.

4.1 Region Content File

Region content file stores all of base geology information, composed of one or more records. To the geological model based on the pole data, the base information includes: the name of area (AreaName), the range of area, model type, model scale, information of control point, numbers of polyhedron in the model, and so on.

4.2 Polyhedron Data Structure

The data record includes the identifier of polyhedron (BodyID), polyhedron attribute (BodyAttris), the volume of polyhedron, the face numbers and identifier of polyhedron (BLF), the identifier of adjacent polyhedron (ABID), and so on.

4.3 Face Data Structure

The data record includes the identifier of face (FaceID), the area of face, the identifier of out circle (CIRCLE), the identifier of adjacent polyhedron (PBID,dNBID), and so on.

4.4 Circle Data Structure

The data structure includes the identifier the circle (CID), the internal face identifier of circle (CIF), the number of edge and arris composed of circle, the identifier of edge and arris (EAID), and so on.

4.5 Triangle Data Structure

The data structure includes the identifier of triangle(TID), the identifier of face(FID), three points of triangle(TPS), the identifier of near triangle, the visible of triangle, the derivative in the point, and so on.

4.6 Arc Data Structure

The data record includes the identifier of arc, the identifier of face, the number of start and end control node or point, the amount number of point or node, and so on.

4.7 Edge and Arris Data Structure

The data record includes the identifier of edge or arris (EID, ArrisID), the identifier of face (FID), the node number of start or end, the visible of edge or arris, and so on.

4.8 Node Data Structure

The data record includes: the identifier of node(NodeID), the identifier of control point (Np), the number of edge and arris connected the node(EANum), the identifier of edge and arris(EAID), the visible of node, and so on.

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms

4.9 Data Structure of Control Point

The data record includes the identifier of control point (NpID), the value of control point, the visible of control point, and so on.

Data structures are connective each other. Connective operations among data are realized via the identifier of data structure. For example, the name of region, the range of region, the scale of model, numbers of polyhedron in the region, and identifier of polyhedrons in the region can be gotten based on the region content file. The data of polyhedron connect the data of face with the identifier of face. The data of face connect the polyhedron data with the identifier of positive, negative face adjacent polyhedron. The data of face connects the data of edge and arris with the identifier of edge and arris. The data edge and arris accesses the data of node via the identifier of node, and so on. Additional, the data of attribute is stored in the region content file and the data of polyhedron, this makes sure of finding the data of attribute according to the identifier of polyhedron and vice versa. This makes it possible to manage the thematic and geometric characteristics parallel. When the change of the identifier of polyhedron happens, the data of attribute is changed also, and vice versa. The internal relationship between the data structure of polyhedron code is shown Figure 6:

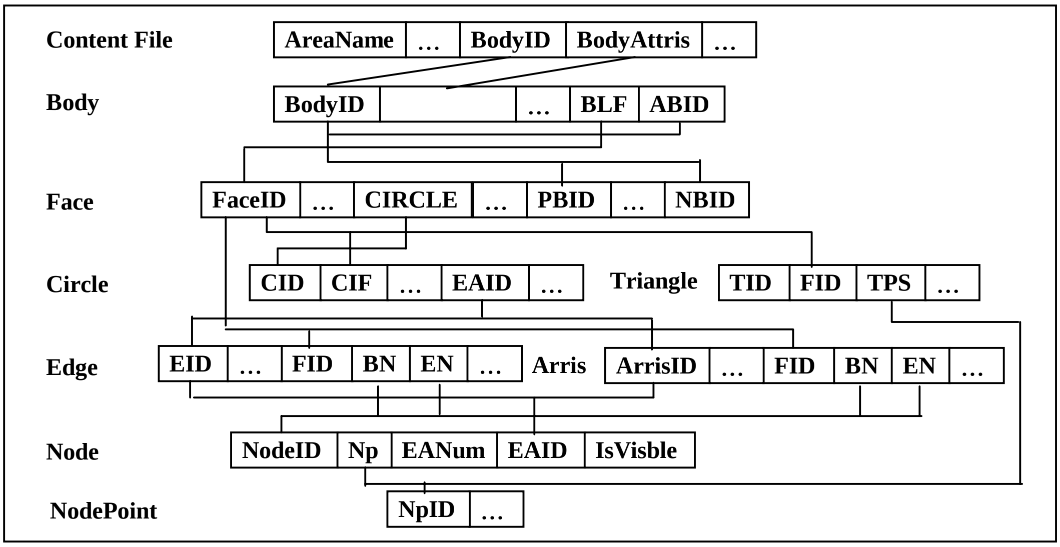

Figure 6 Internal Relationship among the Data

5 CONCLUSION

Geoinformation visualization system integrates common database as the kernel of the system, and specialize in the storage and management of all types of data with unique 3-D visualization and geoscience's analysis.

- Concerning the geological domain, the point, line, face, and body spatial geometric objects are abstractly regarded as body objects. There is only one geometric class--polyhedron in the 3D Vector structure model.
- Polyhedron Code uses high geometric precise to represent the position of 3D geological entity in the Euclidean-space. we apply the relational database in Visual system. Using the relational database model, data is stored as a collection of tables with common fields used to link related records.

REFERENCE

1. E.Bruce MacDougall, Exploratory Analysis, Dynamic Statistical Visualization, and Geogphic Information Systems, Cartography and Geographic Information Systems, Vol.19, No.4, 1992, pp.237-246
2. Jack R. Lousma, Rising to the Challenge: The Role of the Information Sciences, Photogrammetric Engineering & Remote Sensing, Vol.59, No.6, June 1993, pp.957-959
3. Yan Xia, Study on 3D Vector Topoloyg Geological Model and Geoinformatics Visualization System, Paper of Ph.D of China University of Mining and Technology, 1997
4. Rongxing Li, Generation of Geometric Representations of 3D Objects in CAD/CAM by Digital Photogrammetry, ISPRS Journal of Photogrammetry and Remote Sensing, Vol.48 No.5 pp.2-11, 1993

Downloaded from SPIE Digital Library on 27 Nov 2011 to 131.227.157.170. Terms of Use: http://spiedl.org/terms