Index Structure for the Multi-scale Representation of Multi-dimensional Spatial Data in WebGIS*

ZhihanLV[1], GeCHEN[1], RuinaMA[2], YongHAN[1], TaoZhang[1]
(1. College of information Science and Engineering，Ocean University of China，Qingdao 266100，China)
(2. Institut de Recherches en Communications et Cybernétique de Nantes, UMR CNRS 6597, Nantes, France)

(*This research is partly supported by the Natural Science Foundation of China (Project No.: 40730530) and the National High Technology Program of China (Project No.:2008AA09A404))

Abstract: To solve the problem that existing data structure cannot support the multi-scale representation of multi-dimensional spatial data in Web Geographic Information System (WebGIS), a modified data structure has been put forward: (1) getting the main tree from the deformation of the index structure of region quadtree partitioned on the basis of the rule of pyramid structure; (2) possessing the sub-tree structure supporting the overlap of multi-dimensional spatial data; (3)reflecting the changes in spatial resolution using the depth of the tree; (4)all the nodes of the tree are the container of spatial objects. The necessity of generating the index is analyzed which is described. The algorithm for data generation of index structure, the support to the multi-dimensional data and the query process is discussed. For the same data source, some comparative experiments are provided using this structure and layer, and the result shows that this index method can represent and search massive multi-dimensional spatial data effectively in WebGIS. The structure has been used in Shanghai multi-dimensional WebGIS system.

Key words: Network Spatial Index; Multi-dimensional data; VR; WebGIS; 3D

Introduction

In the network information age, with the rapid development of the computer technology, Web Geographic Information System (WebGIS) has also been developing rapidly. It is the integrated production of cartography and computer networks. The application scale and demand are increasing. WebGIS, known as web-based geographic information system, is based on the Web technology under a special environment. In the Internet, it can storage, deal with and analysis spatial information, also can display and apply it[1]. To solve the different application requirements and analysis needs of geographic data, multi-dimensional data processing and multi-scale expressed online have now become urgent and practical problems of geographic information science to solve [2].

There are many issues involved in the multi-dimensional data processing and multi-scale expression, but they can be summed up in four aspects：The first is multi-dimensional data into a single-dimensional problem, now having more mature technical means in industrial sector; second is how to split the division of geographical space, that is to distinguish between multi-scale sequence of sub-space [3], which is more representative is the research from the perspective of spatial cognition [4-5]. Third, is how to generate the sub-space sequences, while dynamically creating and maintaining spatial relationships between them, namely, spatial multi-scale data processing problems, many scholars have studied it from the perspective of this cartographic generalization, and have achieved some initial results , such as the processing model based on Delaunay graph [6] and so on. Finally, is how to manage the generated sequence of sub-space, namely, the establishment of multi-dimensional data indexing structure. This problem increasingly has been paid attention. It not only has great influence for multi-scale representation of spatial data in the speed, but also determines the speed of its query and analysis. The basic idea to solve this problem is to improve the existing spatial data indexing method to fit the multi-dimensional expression of spatial data in network environment. In the traditional one-dimensional spatial data index, more representative index structure is based on quadtree multi-scale representation of spatial data [7]. To solve this problem, we need to create a new expression of multi-dimensional spatial data index structure for the network environment.

1. Analysis of multi-scale representation and search of multi-dimensional data in WebGIS

Based on spatial entity's position and shape or some kind of spatial relation between the spatial entity, spatial index will order a data structure that contains summary information of spatial entities such as the identity of the object, bounding box and the pointer to spatial entity data. To establish spatial index, we can divide the spatial objects by the relationship between them, and establish a space to space object maps, describe the data location information stored in media, access to space object based on partition block. Spatial index can improve the efficiency of the system on data acquisition, not only improve the display speed, but also provide appropriate means for spatial data searching and help improve search efficiency.

WebGIS spatial index is proposed by the two reasons. Since the data is stored on the server disk, if the location of the data on the disk is not recorded and organized, for each query a server must get the entire data file, which access the server disk will seriously affect the cost of efficiency of the system. Therefore, the system designer must record and organize the location of the data on disk, using a number of calculations in memory to replace disk aimlessly for visits in order to improve the efficiency of the system, especially the GIS involved in all kinds of mass of complex data, the index is crucial for the efficiency of processing the data. The second is in multi-dimensional. WebGIS network environment, multi-dimensional geographic data makes R-tree index not apply. In the traditional single-alone space index, R-tree index makes excellent performance, and it has been widely used in the prototype research and commercial space database system such as ORACLE. It is one of the most popular multi-dimensional index structures in stand-alone platform at present. However, in distributed massive spatial data management process in WebGIS, to build R-tree is time-consuming, inefficient in data- update and from the global index there are maintenance difficulties. So it is needed to study a special space index to adapt to the multi-dimensional characteristics of WebGIS.

The methods to establish the network spatial index in accordance with different methods of spatial division is divided into the regular division of the region and object based division. The regular division of the region is to split geographic regions of space in accordance with a certain rules or semi-regular way. Split units and space elements are indirectly associated. The geometry of space elements may be divided into multiple adjacent units. Meanwhile, the description of space-time elements remain intact, spatial index unit only storages the reference information about space elements' addresses. The regular division includes the regional quadtree, BSP tree, octree, KD tree and R-tree series.

Description of the region quadtree index based on the regular division:

(1) Index Theory

In the quadtree spatial index mechanism which based on a fixed grid spatial division, the scope of 2D space is divided into a series of the same size chessboard-shaped rectangle, and thus the establishment of the pyramid structure of N-level quadtree. In the quadtree, the space element of identity are recorded in each leaf node covered by its external bounding box, when the four brother nodes of the same father must all record the identity of space elements, only let the father node record it, and push this rule to the upper forward。

(2) Index Analysis

The grid-based quadtree index method and the composition of grid Index are somewhat similar, both in the form of many to many. But it has effectively reduced the duplicate records of large space elements across multiple grid nodes. And the insertion and deletion of space elements in this mechanism are relatively simple. We just need to record or remove their logo in its coverage of the leaf nodes and the father and ancestors nodes according the above rules. There is no the same as the R-tree complex and time-consuming to split and re-insert operation. Its query styles are relatively simple.

As can be seen from the above analysis, now popular in stand-alone environment, such as R-tree index can not meet the multi-dimensional data distribution network for efficiency requirements, and the current region quadtree Indexing, while suited to network data release, but does not support multi-dimensional data, so it needs to design a index structure for the multi-scale representation of multi-dimensional spatial data---MDDRQ-Tree (Multi Dimension Data Region quadtree).

2. MDDRQ-Tree

MDDRQ-Tree is a deformation of the regular division quadtree index of the region based on pyramid in network environment, which supports online multi-source, multi-dimensional data fusion and multi-scale representation. With the introduction of the pyramid structure of traditional rule-based segmentation of the region quadtree index in comparison, MDDRQ-Tree added a multi-dimensional overlay layer to express multi-dimensional data. Figure 1 is diagram of the MDDRQ-Tree structure , as well as a unit linkages with other units。

2.1 Data structure of MDDRQ-Tree

The data structure of MDDRQ-Tree and the relationship between main-tree and dimensional overlap sub-tree in the structure can be described by two means.

2.1.1 “Father and son" and " property " as an inheritance relationship description

This describes modeled the classic means of “father and son” expression of the quadtree structure. MDDRQ-Tree structure consists of the neighborhood relations within layers ,"father and son" relationship between the layers, and their "inheritance" relationship, in which the main tree is divided into "father" and "son", overlapping sub-tree as their "property". With the increase of the map resolution, the family is expanding, and the possession of the “property” by every generation family member gradually becomes uneven. Defined as follows:

Definition 1. Each unit that is not in the bottom layer has a group of "son", they can be seen as the input of the unit;

Definition 2. Each unit that is not on the top layer has a group of "father";

Definition 3. In the same layer, each unit has a set of "brothers", also known as neighbors;

Definition 4. Since the expansion of the family, the property originally belonging to "father" is re-classified by a group of his son. The other hand, the inherited property of each “son” is part of the division of the father’s property;

By the above definition of the division of MDDRQ-Tree property, the phenomenon is as follows:

Since each "son" in the peer-brothers has different weights, property can not strictly be divided equally; ② After several rounds of the distribution of property , some of the "sons" are no longer assigned to the property; ③ No matter how many rounds of re-distribution of property, the final the property is still all belong to this family, that is, all nodes in possession of the property are the offspring of the property source father .

Figure 1: the right expresses MDDRQ-Tree with the means of "father and son" and "property" of inheritance , pink region describes the property section enjoyed by the whole family.

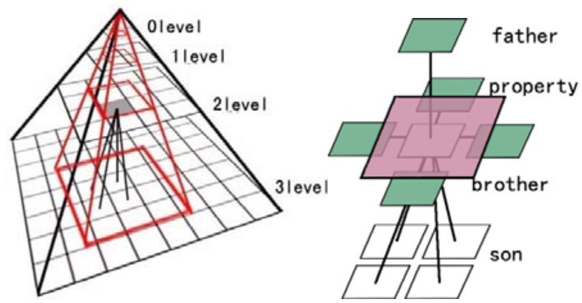

Figure1 . MDDRQ-Tree structure

2.1.2 Graph Theory Description

Describe MDDRQ-Tree Structure with diagrams. The left of picture 1 shows expression of multiple layers overlapping pyramid.

The main tree G= (V, E) consists of the vertex set V and edge set E. Each pair of vertices (v1, v2)∈V×V, has an edge e∈E to link them. The vertices v1 and v2 are called the terminal vertex of e.

Overlapping sub-tree I= (U, F) consists of the vertex set U and edge set of F. Each pair of vertices (u1, u2)∈U×U, has an edge f ∈F to link them. The vertices u1 and u2 are called the terminal vertex of f..

As the vertex set V and U are the abstract expressions of the area region , while the U is the abstractions of non-uniform set, the vertices of U have different weights ,and meet U ⊆ V, F ⊆ E.

Describe MDDRQ-Tree Structure with horizontal and vertical diagram.

Definition 1 Each horizontal diagram can be described by a neighborhood map.

A vertex p ∈ V, can be defined by the following formula:

$$H(p) = \{p\} \cup \{q \in V_i \mid (p,q) \in E_i\}$$

The property Vertex m∈U occupied by the Vertex p can be defined by the following formula:

$$H(m) = \{m\} \cup \{n \in U_i \mid (m,n) \in F_i\}$$

Definition 2 Describe the vertical map with a bipartite graph.

Order $G_i = \{(V_i \cup V_{i+1}), E_i\}$ and $E_i \subseteq (V_i \times V_{i+1})$, $I_i = \{(U_i \cup U_{i+1}), F_i\}$ and $F_i \subseteq (U_i \times U_{i+1})$, for $q \in V_i$, its property is as follows:

$$m = ROOM(q) \in U_i$$

Its son set as follows:

$$SON(q) = \{p \in V_{i+1} \mid (p,q) \in E_i\}$$

the property set of its son as follows:

$$SON(m) = ROOM(SON(q)) = \{p \in U_{i+1} \mid (p,q) \in F_i\}$$

Similarly, for a unit $p \in V_{i+1}$, its property is as follows:

$n = ROOM(p) \in U_{i+1}$

Its father set as follows:

$FATHER(p) = \{q \in V_i \mid (p,q) \in E_i\}$

the property set of its father as follows:

$FATHER(n) = ROOM(FATHER(p)) = \{q \in U_i \mid (p,q) \in F_i\}$

The structure of the MDDRQ-Tree that has n layers can be described by n neighborhood maps and n-1 vertical diagrams。

The structure of MDDRQ-Tree can be evaluated with curtail window, which is the deformation of quadtree. Curtail window is a $n \times n$ square window, contact cell of current layer with a set of cell of sub-level. Curtail window of MDDRQ-Tree is $(2 \times 2)/4 = 1$. So in MDDRQ-Tree structure, each unit has only a father, and there is no multi-parent inheritance.

2.2 MDDRQ-Tree index algorithm

First of all, server-side renders vector data, and then spreads the rendered images in real-time to front-side browser to display. In this process, the amount of the vector data determines the display speed of map. We will render tiles with png format pictures of alpha channel, making the rendered tiles data support the overlay layer effect. Then we can ensure picture quality while also minimize the volume of image as much as possible. In MDDRQ-Tree data structure, Level0 divided the map into m * n regions, each region of Level0 splits into four sub-trees ,and the layer in which the sub-trees are is Level1,then the number of rows and columns in map of Level are as follows:

Number of rows: m×2Leveln, Leveln≥0

Number of columns: n×2Leveln, Leveln≥0

Assuming that the extreme values of the map scope are XMin,XMax,YMin,YMax, X offset is Δx, Y offset is Δy; abscissa difference Δx, vertical difference Δy; the method to calculate the tile position(i, j) by the geographical coordinates(x, y) is as follows:

$$i = \left[\sqrt{(Y_{Max} - y)/\Delta y \times n \times 2^{Leveln}}\right]$$

$$j = \left[\sqrt{(x - X_{Min})/\Delta x \times m \times 2^{Leveln}}\right]$$

In MDDRQ-Tree data, zoom, shrink, move, etc. operation, can be moved into a new map of the scene in the map data coordinates of data needed to calculate the location of tile.

The location of the tile needed will be calculated via map coordinates of the data moved in the scene, while MDDRQ-Tree data zoom, shrink, move, etc. operation. There is no need to generate picture repeatedly, just download the new tiles into the visual domain to fill the missing region, then the efficiency of the server rendering and transmission is improved.

In order to display more extensive geographic information in map, we insert 3D landscape map data into 2D vector map using the main tree index of MDDRQ-Tree structure, as a overlay sub-tree layer. Overlay layer structure changes following the move of the main center coordinate of the map , to indicate the 3D area in metro map. And conjunction with Ajax technology [9] makes the 3D region can viewpoint in real-time blanking when the map moving, as shown in Figure 2,when the map is moving, the tiles of MDDRQ-Tree overlay sub-tree are dynamically removed and loaded.

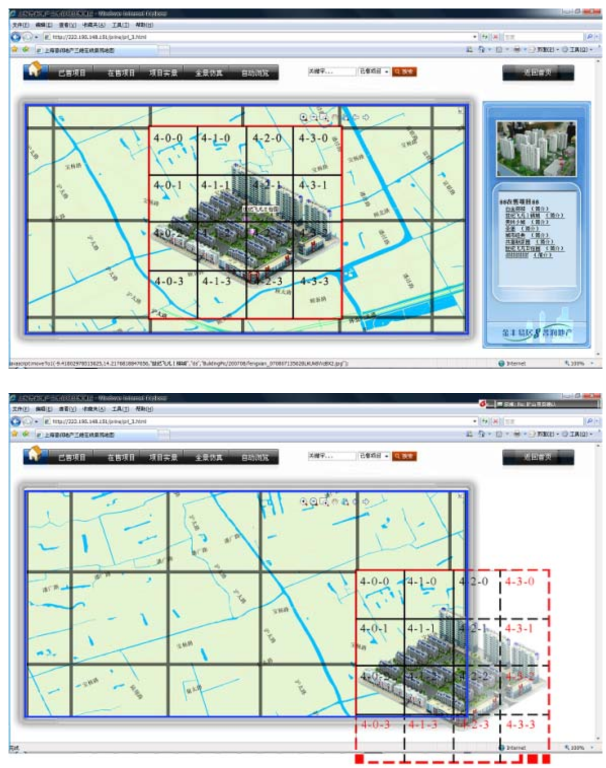

Figure 2 Tiles of MDDRQ-Tree overlapping sub-tree are removed dynamically

2.3 MDDRQ-Tree on the support of multi-dimensional data

Through the realization of rendering 2D vector data into tiles of main tree structure by vector data rendering engine just in time, pre-process three-dimensional data to overlapping sub-tree tiles, and unify them into MDDRQ-Tree space.

2.3.1 MDDRQ-Tree on the support of

two-dimensional vector data

Vector data is based on the rectangular coordinate system, with points, lines, polygons to describe the data model or data structure of the geographic elements. Each geographical element is described by a series of ordered sequence (x, y) coordinate , these elements are combined with attributes 。Taking into account the system's scalability, we designed Geometry model strictly following the OGC OpenGIS Simple Features data specification . Geometry defined the basic geometric shapes in a vector graphics, as well as some simple spatial association analysis function. The most important geometric shapes are point, line and surface, and multi-point, multi-line, multi-faceted, curve, straight line, ring, polygon, etc[10] that derived from point, line and surface , shown in Figure 3.

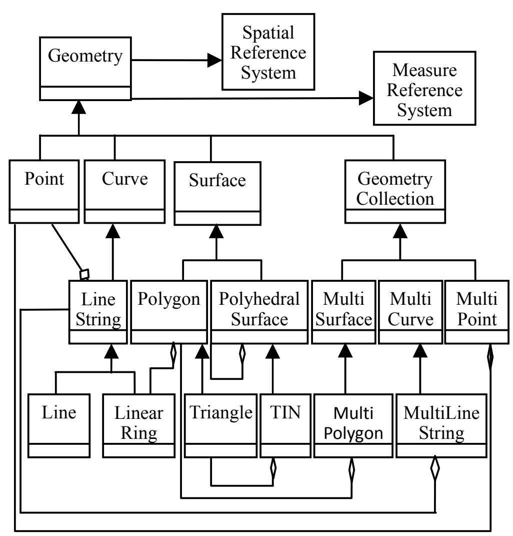

Figure 3 Geometry Model

Layer container is achieved. As a series of collections with the same type of Geometry characteristic, all the features of a layer unify the coordinate system and property field. Design a top container Map scene manager to unify the operation of the entire map, and construct index of all the layers to manage them, forming an inheritance relationship and level management system like Map-->Layer-->Geometry.

Build optional R tree or quadtree indexing mechanism for spatial data, so as to increase the efficiency of spatial data query and analysis. Take each geometric shapes as a record, and store it on the feature dataset, and set an index number for this record providing a unique identifier. Through the quadtree data organization method, we can quickly locate the space record we need to find. Coordinate System is Cartesian space coordinate corresponding window coordinates, having realized the transformation between longitude, latitude and Descartes coordinate system. Transform the longitude and latitude of geometric data eliminated by quadtree into the relative coordinate, then take this as offset to display the geometry to the window. In order to achieve the fusion of multi- scales and multi- resolution vector with raster data, and realize the projection transformation of spatial coordinate system, in this system we do projection transformation from longitude and latitude coordinate to Lambert coordinate that is usually used in china map, making spherical data can be fully manifested into 2D scene. Using Object- Oriented technique and design patterns thought to abstract and layered the functions of each module, and publishing it to a component [11], vector data as a server- side real-time render engine, and provides fine- grained interface for the browser call, such as shown in Figure 4.

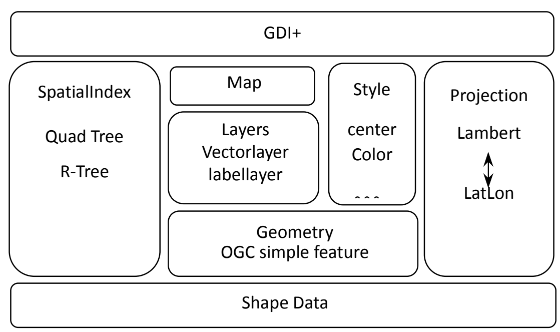

Figure4 Real-time vector rendering engine

2.3.2 Pre-processing the 3D data of MDDRQ-Tree

3D landscape maps and paper maps not only have different storage media, in fact, on the three elements of the map, mathematical laws, map symbols, and cartographic generalization, there are significant differences between the 3D landscape maps and traditional maps [12].

3D landscape map is actually a kind of observation of surface-rich 2.5-D, it is static raster data generated from 3D model rendered on a fixed perspective. According to the orientation of the house in the region we determine the camera angle, in 3D modeling software, set the camera and lighting effects, and produce base maps, then the standard scenario form. We prepare it as a script. The objects that are blocked, or in a secondary position or beyond the horizon can be omitted or simplify.

The point, line and plane city elements information layers of 3D scene are hierarchical productions in the model making. Their superposition forms 3D landscape of the city. The city is the scene of a huge amount of data. So in grasping the effects of the map simulation, we cannot blindly pursue high-precision and reduction of city[13] . In the process of model rendering, using Mental-RAY render through the calculation of the parameters such as the triangle face number of models, the particle sampling number of single pixel and intensity of light emitted and so on, we can realize the scene rendering [14]. Multi-camera rendering images form an overall map by stitching. Based on different scales the overall map is cut into 3D visual map tiles that meet the requirements of the client MDDRQ-Tree index algorithm, then storage them on the server side. When a client browser do the operations such as zoom in or zoom out, 3D landscape map tiles and 2D vector map tiles together do MDDRQ-Tree removing optimization of the data to ensure the efficiency of data transmission。

2.4 The search process of MDDRQ-Tree

MDDRQ-Tree index storages the structure of main tree and overlay tree and the relationship between them with the array, in the tree node, storages the code of each region of space. It is very convenient to access and retrieve them in a network file.

MDDRQ-Tree search algorithms include the main tree search and overlay sub-tree search. The main tree search is, separating evenly the region of the primary tree map, for those objects falling in the specified partition area, index coding and implementation of space object query. Define the bounding box of partition area with vector4(float, float, float, float). The point object of the space can be seen a rectangle of which the 4 coordinates of the bounding box are coincident. The line object of the space can be seen a rectangle using its bounding box as index[15]. Calculating the geocoding of corresponding coordinate, and saving the result as a member of the spatial index in the member variables of spatial object. When you need to retrieve space objects, first filter all segmentation areas intersecting spatial object bounding box, then traverse the index rectangle of screen results, by the judgment whether the 4 encoding is located at the specified index rectangle or not to determine the need of the space object data. Overlay sub-tree search only contains the search of the point object. First of all, we need to search out its corresponding main tree, and traverse the main trees in the area of overlay sub- tree tiles, then collect the tiles that contain the destination point object ,and finally, the data set we get is the desired results.

3. Experimental results and conclusions

The experimental platform this paper used is Intel Core2 Duo 2.53GHz, 2G CPU, GeForce 8600 GT graphics card, 256M video memory, Windows XP operating system.

Based on the data in figure 3, compare MDDRQ-Tree based spatial data index structures with structure of layer expression in display performance through the experiment. Structure of layer expression is a common method, that is using layer structure superpose 2D data on the map in the corresponding location directly. Realize multi- dimensional display of spatial data of 1:500 to 1:2000 in order. Table 1 embodies under the different display - scale, two indexes have different time - consuming.

From the vertical, multi-dimensional spatial index using MDDRQ-Tree methods is not more time-consuming due to increased resolution. From Lateral from the table to compare the results, in the same resolution, the higher the resolution is, the greater the gap of the time-consuming between MDDRQ-Tree index method and layer expression.

Scale	MDDRQ-Tree time-consuming/ms	Layer expressions time-consuming/ms
1：2000(1st time)	63	78
1：2000(2nd time)	62	78
1：2000(3rd time)	59	94
1：2000(4th time)	61	78
1：1000(1st time)	64	266
1：1000(2nd time)	61	265
1：1000(3rd time)	61	281
1：1000(4th time)	64	265
1： 500(1st time)	64	1015
1： 500(2nd time)	63	1016
1： 500(3rd time)	63	1031
1： 500(4th time)	62	1000

Table 1 Comparison of indexes time – consuming

From the experimental results we can see, MDDRQ-Tree not only can meet the network multi-dimensional spatial data expression requirements in the display effect, but also has greater efficiency compared with the traditional layers expression.

In the experimental platform, we are using VS 2005

(C #) Asp.net2.0 development environment, SQL Server 2005 databases, Microsoft IIS 6.0, as well as JavaScript, Ajax techniques to build separate front and exchange data between front and back, then achieve the function of the system. Taking Shanghai as an example, the performance of the system has been tested. Vector data used the shape data, the area is 6000 square kilometers, scale is 1:10000, the region of 3D Landscape Map is Fengxian, with an area of 25 square kilometers. Use 3DS Max modeling and render it to the 3D landscape map by a rendering script supported by the platform, and then pre-process the data to MDDRQ-Tree overlay sub-tree tiles; the data includes 20 3D POI buildings information, 20 3D cell scene models and 10 indoor and outdoor panorama. Shanghai 3D WebGIS example is shown in Figure 5, the user do not need to download additional plug-ins. Experimental results show that it have achieved the desired effect. Part of Fengxian District, Shanghai, and the software interface are shown in Figure 6.

Figure 5 3D WebGIS of Shanghai, China

Figure 6 Part of Fengxian District, Shanghai, China

4 References

[1] Peng ZR, MH. Tspu, Internet GIS-Distributed GIServices for the Internet and wireless networks,2003 Wiley.

[2] Han,J. and Y. Fu, Mining multiple-level association rules in large databases. IEEE Transactions on Knowledge and Data Engineering,11(5),1-8(1999).

[3] Brakatsoulas S, Pfoser D. and Theodorides Y..Revisiting R-Tree construction principles. Technical report, Computer Technology Institute, Patras, Greece, 2002.

[4] Vange not Christelle.Multi-representation inspatial database using the MADS conceptual model. Proceedings of the International Cartographic Association Workshopon Generalization and Multi-Scale Representation.Leicester.England.2004:337-342

[5] Mark. David M. Freksa Christianetal. Cognitive models of geographical space. Geographical Information

Science.1999.13(8):747-774

[6] Peng Hu. QiQing- Wen. LiuZhao- Li. Progress in studies on automated generalization of spatial point cluster. IEEE International Geoscience and Remote Sensing Symposium. 2004.13(8):2841-2844

[7] Li Ai-Qing . Research on the seamless spatial database and its multi-scaling representation and processing[D]. Wuhan: Wuhan University, 2001(in Chinese)

[8] Guttman A.R-tree: A dynamic index structure for spatial search. Proceedings of the ACM SIGMOD International Conference on Management of Data.Boston.US.1984:47 57

[9] SayarA. Pierce M. Fox G. Integrating AJAX Approach into GIS Visualization Web services[J].IEEE Computer Societv.2oo6(2) 169~170.

[10] OGC.OpenGIS Simple Features Specification 06-103r3_Candidate_Implementation_Specification_for_Geographic_Information_-_Simple_feature_access_-_Part_1_Common_Architecture_v1.2.0

[11] Yudong Luo, Weijun Chen. Research and Development of Com-GIS, COMPUTER ERA ,2004(2)

[12]Nobuyuki Bannai, Robert B. Fisher, Alexander Agathos. Multiple color texture map fusion for 3D models[J]. Pattern Recognition Letters ,2007,(28),748～758.

[13] Qing Zhu, Yurong Gao, Yongjun Wei,Duo Huang. Design of 3D Models for 3D GIS [J]. GEOMATICS AND INFORMATION SCIENCE OF WUHAN UNIVERSITY, 2003, 28(3):283~287

[14] B. de Cambray. Three-Dimensional Modelling in a Geographical Database. 11th Int. Symp. on Computer-Assisted Cartography, pp. 338--347, 1993.

[15] L. Brun and W. Kropatsch. Introduction to combinatorial pyramids. In:R.Klette,G.Bertrand and A.Imiya,Editors.Digital and Image Geometry LNCS vol.2243,Springer,Berlin (2001),pp.108–127.