THE STUDY ON RAPID VISUALIZATION
OF GLOBAL MASSIVE SPATIAL DATASET PROCESSING

Hujingbo [a], Xionghanjiang [a], Zhangyongzhi [b], Zhaofusheng [c]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing , Wuhan university,
430079 - raining1219@163.com, xionghanjiang@163.com;
[b] Bengbu Tank Institute, 233000 - lyzyz@tom.com;
[c] School of Info-physics and Geomatics Engineering, Zhongnan University - zhaofusheng204@sohu.com.

ThS-4: HIGH PERFORMANCE VISUALIZATION

KEY WORDS: Digital earth, Visualization, Block, Image, Multi-resolution, Three-dimensional.

ABSTRACT:

Virtual Digital Earth or the system of terrain visualization belong to the system of data-intensive applications, the data which applied to the system is from hundreds of MB to TB. If the virtual systems of the network model achieves the rapid real-time visualization of the mass data, the key is the organization of data. In view of this, the paper build a multi-resolution tile pyramid as the global data to study, and it provides the block-handling algorithm. The most fundamental aim that we adopt this strategy is to reduce the amount of the data which participate in topographic drawing reasonably, thereby we can reduce the time to simplify the complexity of the algorithm.

1 INTRODUCTION

In 1998, U.S. former vice president Gore put forward the concept of "Digital Earth" in the annual conference of the Open GIS Consortium. He believed that Digital Earth was the Virtual Earth with multi-resolution which could be three-dimensional expression and was the basis of the coordinate of the Earth by the mass of data. There is a substantial progress in the process of the global information along with the concept of Digital Earth proposes. People began to pay attention to the study of three-dimensional data instead of the past two-dimensional images. It is the researchful goal of the scientists, which is how to show the real world which we live more truly.

Virtual digital Earth or the system of terrain visualization belong to data-intensive applications, its typical point is the rapid visualization of the mass spatial data. Terrain data for the application of the system is about between several hundreds of MB to GB in general, even to TB. For example, the DEM data of the world's 30 m resolution has been more than 12 TB. Because of the massive characteristics of the geospatial data and the current conditions of graphics hardware restrictions, three-dimensional topographical drawing at the scene could not load all the terrain data to establish the whole terrain scenes once. Therefore, simply relying on the hardware technology is impossible to resolve the problem of massive data processing, in order to achieve the purpose of the rapid global spatial data visualization, it is necessary to combine hardware and software to solve this bottleneck of massive data processing.

2 THE ORGANIZATION AND MANAGEMENT OF GLOBAL MASSIVE DATA

The key of the visualization of the virtual digital earth is the organization of data, the difficulty of it is how to deal with the massive global data.

2.1 Summary of Issues

With the development of surveying technology, the number of geospatial data increases by a series growth. For example, the size the digital orthophoto with 1-meter resolution is 100,000 * 100,000 pixels in 100 * 100 square kilometers area. If each pixel value occupies three bytes (24 color), it needs about 28 GB of disk space. The size of Digital elevation model of four meters resolution is 25,000 x2.5 10,000 elevation points, if the value of each elevation occupies four bytes; it needs approximately 2.4 GB of disk space. If we extracted all the data from the disk to draw the whole terrain, it will produce square of 625 million or triangle of 1.25 billion, and it needs texture drawing for 2.5 billion or 3.75 billion. It is intolerable to bear such a huge operation to the general configuration of the graphics workstations.

In order to resolve this problem, the paper adopts the organizational structure based on the block model of the pyramid and the spatial data indexing mechanism based on global Quadtree. The essence of this approach is to reduce the amount of data which actually deliveries to display card for drawing. Thereby it can reduce the Complexity of the algorithm and improve the efficiency of drawing. Because of the orderliness of the observation by human eyes, it could only receive less information in the more distant the scene in view of the limited-scope. Along with the shorter distance, people could apperceive the more and more growing details. Therefore, it needs to build a model of level of detail (LOD) and gives people a real sense of vision.

2.2 The Strategy of Regular Grid Block

In order to realize three-dimensional visualization of terrain data, it is not very realistic to load all terrain data by one time aimed as massive global spatial data by the computer's capabilities. And it is impossible to show the whole region of three-dimensional scene in a small screen and to be sure that the details do not show distortion. Therefore, we have adopted the

method of block, and we read the data block as a unit for writing. When the observers' eyeshot space (view volume) arrived at a certain position in the program, the program would be automatically catch the corresponding terrain document in the position, and when a block of visual scene leaved the view volume, the program could uninstall the terrain data blocks so as to ensure effective and sufficient use of system resources.

There is no general standard to follow the method of block purely, but the irregular block will directly affect the construction of the index and pyramid building, and if will affect the operation of index and the efficiency of reconstruction. In addition, the data block which is too big or too small will affect the effective performance of the system. If the data block is too much, it may lead to read too much redundant data (it is not the data in the target range). If the data block is too small, although it reduces the redundant data, it will increase the times of addressing disk and read operation. And it is not conducive to save the entire time of reading data, and it will influent the network transmission. Therefore, the key of the data block organizations is the size of the data block, the block size should be conducive to the distribution of storage, and it should be conducive to the building of pyramids, and it is conducive to Quadtree index and so on. To sum up the above objectives, we adopt the rules of regular. The equal block-size is that the wide and height are both equivalent between the block and the blocks; the equivalent rule is a law, rather than a chaotic sequence. The size of the data block unit is the integral power of 2. The size of the data block which is restricted to a 256 x256 pixels is the most suitable for the system of equipment by research, and it could achieve the best performance of the disk to read and write.

We plot out the rectangular block of the topography surface in accordance with the above method and code, and then we establish the grid index. The principle of the grid index is to put the topography surface into a spatial reference and is divided into a consecutive rectangular block and to take the coordinate origin as a reference, and we code those data blocks according to the range. We take the data retrieval and data scheduling according to the scope of the block unit when in use.

2.3 Tile Pyramid Model

Pyramid is a multi-resolution model of level, see Figure 1.
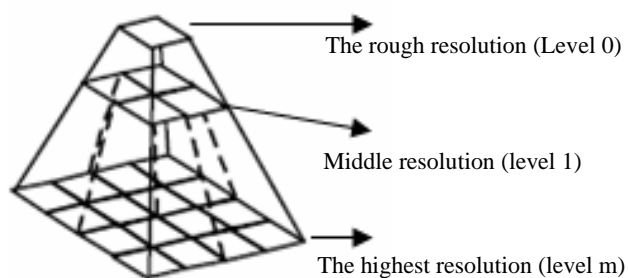

Figure 1. Tile Pyramid Model

When we draw the terrain scenes, we often need different resolution of the geospatial data in different regions in order to increase the speed of show ensuring the accuracy displayed on the premise. That is the low-resolution imagery of satellite and the terrain data is used to describe a wide range of macro-topographical features and the high-resolution imagery of satellite and terrain data is used to describe the landscape features of some areas in detail. To an accurate extent that the pyramid model is a model for the resolution of consecutive

change, but because it is very difficult to build the pyramid in the resolution of consecutive change and it has no sense to do like that, we adopt the approach of the general rate to construct the pyramid. The resolution becomes more and more coarse from the bottom to the top of the pyramid, but the scope is the same. Generally speaking, the creation of terrain database is usually take 2 as the ratio of the resolution of the pyramid, that is we assume A0 to be resolution of the pyramid of the 0 level and Am to be resolution of the pyramid of the m level:

$$Am=A0*2m \qquad (1)$$

Generally speaking, the data sources provide the resolution between the various data isn't smooth, the resolution span between the neighboring data level may be very large or may be very small. This is not only to cause trouble to the adaptive changes in the calculation of drawing data resolution according to the location of the visual perspective as visualization, but also the appearance of that drawing the picture of the transition is not smooth will impact the visual effects when the resolution data converts. Therefore, the resolution of the adjacent data level takes 2 to be the fixed value. If the data source could not provide the middle resolution of the data, the high-precision resolution of the data will be obtained by resampling. For example, data sources can provide the accuracy of the topographic data is 1 m and 10m resolution, and we need to build the pyramid of five levels for five-resolution data level of the establishment of 1 m, 2m, 4m, 8m, 16m. Because there are not these data for the data level of 2 m, 4m , 8m, 16m, we sample the resolution of 2 m, 4m, 8m through the resolution of 1 m, and we sample the resolution of 16m through the resolution of 10m.

The rank number of the original terrain data which is at the bottom of the rank number of the pyramid data structure stores the bottom terrain data, then we set up the bottom level to be the current data level of the pyramid structure. The rank number of the current level divides 2 is the rank number of the upper level of the current level. We resample the terrain data of the current level according to the relationship of the location to get the terrain data of the upper level, and then we store it in the database. We set up the upper level of the current level to be the current level, and then we repeat the above operations until the end of the top-level data storage.

Although the pyramid model increases the space of the data storage, it can reduce the total time of completing topographic drawing. Tiles pyramid of the block model will also be able to further reduce the volume of the data access and improve the efficiency of the implementation of input and output so as to enhance the system's overall performance. When the size of the window for showing the terrain is fixed, it could keep the volume of the data access unchanged to use tiles pyramid model. The characteristic of tile pyramid model is very important to the real-time visualization of massive terrain.

2.4 The mechanism of the global linear Quadtree index

Quadtree is a level of data structure and its basic thinking is: first of all, the eyeable terrain is separated into four equal-size rectangular block, if it examines that all the rectangular grid has reached the required pixels values, it will do not need to continue to separate, otherwise this rectangular block is separated into four equal-size rectangular block to be further small, this recursive division continue doing until all of the

rectangular grid of the child block reach the required block size value. See Figure 2.
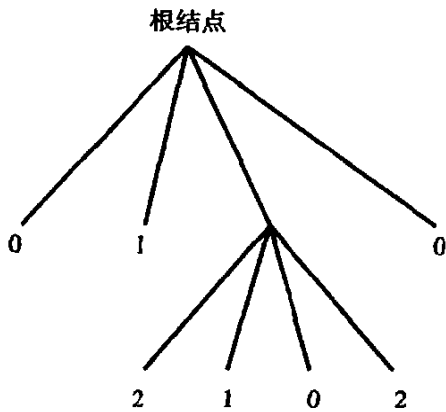

Figure 2   The Structure of The Linear Quadtree

First of all, when we use the structure of quadtree to show the terrain, the terrain data should be resampled by the rule of the equal space between grids of the different accuracy. The result of sampling constructs level nodes of quadtree. For any neighboring level in quadtree, the accuracy of sample in the upper level is a second of the next level. In particular, there are four child nodes in a non-arbitrary end node of the tree, and the sampling area of the child nodes divides the sampling area of the sub-regional node into four equal areas. Each node in quadtree covers a rectangular region corresponding to the terrain. The sampling points of the upper nodes are less and have a higher efficiency of drawing to show the terrain. But they have the lower resolution and the error of the terrain expression is large. The bottom topography have the high resolution and the small error, but it draws with low efficiency. There are three steps to build the linear Quadtree index in the pyramid built: (1) logic block (2) node coding; (3) physical block.

(1)   Logic block
It ordains that the dipartition of block begins at the left bottom corner of the terrain data corresponding to the construction of the pyramid tiles, and it carries out from left to right and from bottom to top by turns. At the same time it ordains that the level coding of quadtree brings into correspondence with the coding of the pyramid. That is, the bottom of quadtree brings into correspondence with the bottom of the pyramid. See Figure 3.

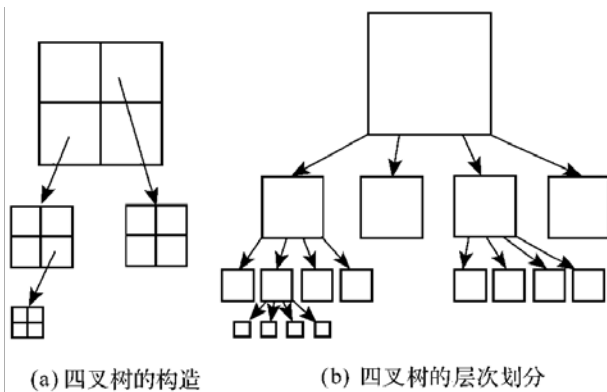(a) 四叉树的构造   (b) 四叉树的层次划分

Figure 3 the dipartition of the global quadtree grid

We set up (ix, iy) as pixel coordinates and is as the size of tiles, and we make io as the overlap of the adjacent tiles, and all of them make pixel to be the unit. We set up (tx, ty) to be the tile coordinates as a unit of block, and we take l to be the level number. If the tiles coordinates (tx, ty) are known, the pixel coordinates of the   left bottom corner of the tiles (ixlb, iylb) are:

$$ixlb = tx \times is \qquad (2)$$
$$iylb = ty \times is \qquad (3)$$

The pixel coordinates of the right top corner of the tile (ixrt, iyrt) are:

$$ixrt = (tx + 1) \times is + io - 1 \qquad (4)$$
$$iyrt = (ty + 1) \times is + io - 1 \qquad (5)$$

If the pixel coordinates (ix, iy) are known, the coordinates of pixels which belongs to the tiles are:

$$tx = ix / is \qquad (6)$$
$$ty = iy / is \qquad (7)$$

By the size of the matrix rows and columns of the pixel and the size of tiles you can calculate the rows and columns of the tile matrix, then the logic tiles will generate from left to right, bottom to top, the logic tile is unique identified by(ixlb, iylb), (ixrt, iyrt), (tx, ty) and l.

(2)   Node Coding
With the assumption that it make one-dimensional array to store the tiles index, the ordering of tiles starts at bottom and carries on from left to right, bottom to top. The offset of tiles in the array is nodes coding. In order to extract tiles (tx, ty, l), we must calculate its offset. We adopt a one-dimensional array to store the starting offset on each tiles floor and set it as osl. If the columns of the lth floor tiles matrix are tcl, and the offset of tiles (tx, ty, l) offset are:

$$offset = ty \times tcl + tx + osl \qquad (8)$$

(3) physical block
It takes the terrain data to be the physical block on the basis of logic block to generate the child block of topographic data. You can fill the invalid value of pixels in the redundant parts on the right side and the top side of the tile. After completing the physical block, you can store them by the number of the tiles.

3   DISCUSSION AND CONCLUSION

Above all, we introduce A 3D-Digital earth data processing system, and it takes the advantage in processing the global data. It uses the technology of data blocking, spatial index, seamless data and the method of multi-resolution pyramid organization to create a multi-scale and multi-resolution pyramid structure. And the data processing capacity of system is up to more than 100 GB to TB-class.

3.1 data preparation

The coordinates we most often use includes the Beijing 54 coordinates \Xi'an 80 coordinates and WGS84 coordinates, the data which our system deals with must be the WGS84 coordinates, if they are not in this coordinate system, and we need for projection conversion.

3.2 the result of the experiment

Figure 4 shows the management system of the multi-resolution DEM data and the image data. The system is based on the organization, management and visualization of the global massive seamless spatial data based on the environment of the network. It carries out that any person, any time and any where to observe any corner of the earth by any height or any angle through the high efficient organization, management and visualization of the global massive image data, topographic data and the data of three-dimensional city models via the Internet.
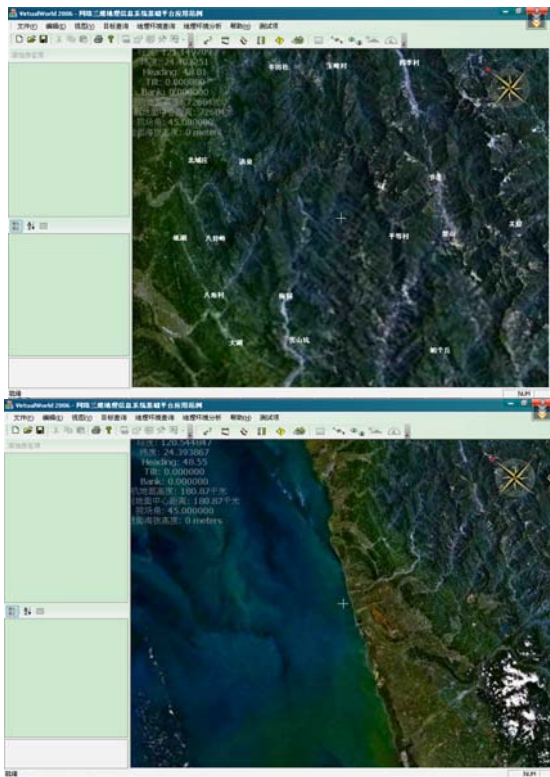

Figure 4   Multi-resolution show of the multi-source imaging data

REFERENCES

1. Liqiang Zhang, Key Technologies for 3D Digital Earth Development [D], Doctor Degree Papers, Chinese Sciences Academy of Remote Sensing Application Technology Institute, June, 2004.

2. Xueqing Deng, Research of Service Architecture and Algorithms for Grid Spatial Data, Doctor Degree Papers of Surveying and Mapping, Information Engineering University, April, 2003.

3. Zhanqiang Chang, Lixin Wu, the Compression of Grid DEM Data Based on Wavelet Transform and Mixed Entropy Coding in the Mountains[J], Geography and Geo - Information Science Vol. 20, No. 1, January 2004.

4. Wei Luo，the Research on Infinite Scale Visualization of Mass DEM Data, Wuhan University Master Papers, may 2006.

Appendix

Each level revolution of pyramid: (WGS-84 coordinate，unit：degree, e.g. image data)

1:     0.0703125   18 degree/256
2:     0.03515625   9 degree /256
3:     0.017578125   4.5 degree /256
4:     0.0087890625   continue like that
5:     0.00439453125
6:     0.002197265625
7:     0.0010986328125
8:     0.00054931640625
9:     0.000274658203125
10:    0.0001373291015625
11:    0.00006866455078125
12:    0.0000034332275390625
13:    0.0000171661376953125
14:    0.00000858306884765625
15:    0.0000042915344423828125
16:    0.000002145767219140625
17:    0.00000107288360595703125
18:    0.0000005364418029785515625
…………………………………and so on