QUAD TREE SPATIAL SPECTRA GUIDE: A FAST SPATIAL HEURISTIC SEARCH IN A LARGE GIS

Zi-Tan Chen
Donna Peuquet

Geography Dept , Univ of California,
Santa Barbara, CA 93106, U.S.A

ABSTRACT

An approach for improving spatial search efficiency in large-scale GIS is presented in this paper. This improvement is achieved by a spatial heuristic search instead of the conventional blind search in GIS. This approach is a practical application of artificial intelligence for solving the difficulties caused by large-scale GIS, which spends excessively long CPU time for many search operations on huge volume data. The traditional spatial search can be categorized as blind search, which wastes many efforts on those impossible regions This research increases the spatial search efficiency by a heuristic search strategy, which finds the candidate areas by utilizing spatial knowledges. The spatial knowledges, required by the heuristic search, are the quad tree spatial spectra(QTSS) trees of spatial data and goal objects A forward up-down strategy controls the search procedure. The interpretation algorithms among the QTSS are developed as the search rules. The test results are shown in this paper.

INTRODUCTION

The requirements for geographic information systems are increasing rapidly. Likewise, the contrast between huge volume data and limited computer sources is also increasing. For large countries such as the USA, China and Canada, size is a real problem. One of the causes of this problem is the inherent limitation of algorithm efficiency. Any practical GIS has to use some 'slow' algorithms because there is not always a better alternative.

After these expensive large computer GIS systems are established, their manipulation efficiency is a serious problem. For practically applying GIS, it is better that the computer system not only can do geographic data storage and simple manipulations, but also that the system is smarter --- we hope GIS has some degree of artificial intelligence(Smith, 1984).

An important concept of artificial intelligence is the utility of human being's knowledges, especially those professional knowledges of the experts(Nilsson, 1980). The heuristic search is a way to use these knowledges(Pearl, 1984). A GIS is a spatial data base. For improving the spatial search efficiency, the heuristic search must be implemented not only on symbolic data, but also on spatial data.

This approach focuses on the spatial location heuristic search based on spatial knowledge extraction and its applications. The geographic data are represented as quad trees. The spatial knowledges are extracted as quad tree spatial spectra (QTSS). (Chen, 1984). Then, a spatial heuristic search control strategy is implemented based on the interpretation of QTSS knowledges.

Those GIS with the capabilities of location heuristic search are more intelligent. The GIS can pick up the higher potential candidate regions just using these compact knowledges. The complex algorithms are only executed on the candidates. Thus, much of the CPU time can be avoided.

Suppose the CPU time for implementing an expensive algorithm for the whole area S is TA. Then the CPU time for implementing at a partial area S/N is assumed to be TA/N. The CPU time for finding the candidate area is TC, while a heuristic search and implement time TH is:

$$TH = TC + TA/N \tag{1}$$

If TC is much less than TA, and N is not small, TH will be much less than TA.

QTSS TO REPRESENT SPATIAL DISTRIBUTION KNOWLEDGES

The principles for evaluating spatial knowledges

The selection from different kind of spatial distribution knowledges to be used in GIS should consider the following principles:

(1) These knowledges include richer spatial distribution information;
(2) the storage of these knowledges takes small space;
(3) various interpretation algorithms can be easily made from these spatial knowledges and they are efficient;
(4) the generation of these knowledge is efficient.

QTSS definition

The concepts of quad tree spatial spectra QTSS have been proposed(Chen, 1984). First, a raster image converts its quad tree representation A $2^N * 2^N$ binary raster image can be represented by its quad tree by successively subdividing it into four quadrants until all pixels within a quadrant are homogeneous(Klinger and Dyer, 1976). The number of black nodes at each tree level constitutes a set. The length of this set is the level number N. For instance, the evergreen forest at Black Hill area, South Dakota, has this QTSS:

(0 0 0 7 117 452 1354 3299 7765 14997 29893)

This QTSS has a length of 11. Such a set is defined as a quad tree spatial spectrum(QTSS) of this spatial phenomenon for this area(Chen, 1984). To judge QTSS based on the previous principles, the QTSS can be generated much faster than those conventional transformations such as FFT or FWT, and they store more compactly. QTSS do not contain complete spatial information, but, nevertheless, they do include a great deal of statistical spatial information, as well as some contextual information. For example, an interpretative function, called as "QAREA", calculates the area of this phenomenon at a quad tree node by the formula:

$$QAREA(QTSS) = \sum_{l=0}^{L} \frac{QTSS(l)}{4^l} \tag{2}$$

where l is the level sequence and L is the length of QTSS. Many interpretations of the spatial distribution properties based on QTSS are possible.

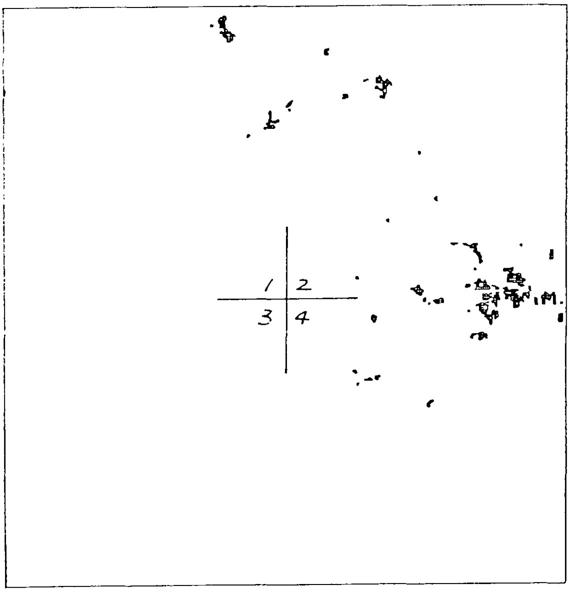
Elementary phenomena represented by QTSS

There are two kind of geographic data in GIS. One is elementary phenomena, which has been input directly from data sources, such as lakes or railways They can be stored in GIS as one simple binary image. Another kind of data is complex phenomena, which consists of several elementary phenomena. A phenomenon can be either elementary or complex depending on the GIS in question. For example, the evergreen forest in a rough GIS of land-use might be an elementary data, but it would be a complex phenomenon in a thematic forest GIS.

The QTSS can describe the spatial distribution properties of an elementary phenomenon, such as density, clustering and approximate location.

Figure 1 shows the quad tree image of the residential areas at Black Hill. The Rapid City is at the middle east of the image. It is an elementary phenomenon in the GIS. Table 1 shows its QTSS at the root (whole image) and its four quadrants. The QTSS at the root shows that some residential areas exist, but the total size of these residential areas only is small fiction of the whole area. Also, they are dispersed. From the QTSSs for four quadrants, it can be seen that most of the residential areas are in quadrant 2 and 4, while a few are in quadrant 1 and none is in quadrant 3.

Figure 1. Quad tree representation of the residential area at the Black Hill



THE ARCHITECTURE OF GIS

The GIS includes three parts· a data base, a knowledge base, and a search module.

A spatial data base

All geographic data are stored ın this data base They are original spatial data, inputed from maps, classified ımages, etc The data structure representıng these spatial data is quad trees All data are registered at a same square area Each elementary geographic phenomenon is converted into quad tree file (Peuquet, 1984).

A spatial knowledge base

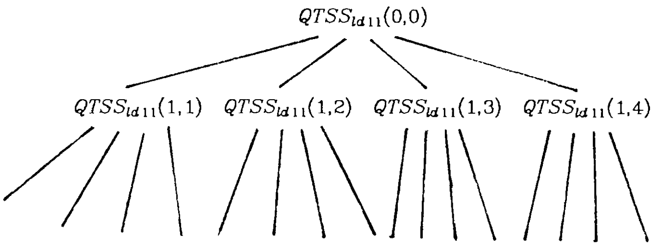
The spatial distrıbution knowledges are stored in this part. These knowledges are the QTSS trees, extracted from these origınal spatial data. First, for each node at the top several levels of a quad tree, one QTSS ıs generated from the original data. Table 1 shows five QTSS of residential area, one ıs in the root, other four are on four quadrants, respectively.

Table 1. An example of five QTSS
of residentıal area (land use code ld11)

$$QTSS_{ld11}(0,\ 0) = (\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 82\ 472\ 1456\ 3311\ )$$
$$QTSS_{ld11}(1,\ 1) = (\ 0\ 0\ 0\ 0\ 0\ 0\ 6\ 64\ 252\ 525\ )$$
$$QTSS_{ld11}(1,\ 2) = (\ 0\ 0\ 0\ 0\ 0\ 1\ 42\ 214\ 692\ 1729\ )$$
$$QTSS_{ld11}(1,\ 3) = (\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ )$$
$$QTSS_{ld11}(1,\ 4) = (\ 0\ 0\ 0\ 0\ 0\ 1\ 34\ 194\ 512\ 1057\ )$$

Where the $QTSS_p$(L, N) is the QTSS of phenomenon P of the N_{th} node at the L level. These separate QTSS can be organized ınto a tree structure called as QTSS tree Fıgure 2 shows a QTSS tree of the same data

Fıgure 2. The structure of a QTSS tree
(only the top two levels of the QTSS tree is showing here)



In the test GIS, an image has 2048*2048 pixels. Only the top four levels of the QTSS tree have been stored in this GIS. The QTSS at root has a length of 11 It means that the smallest candidate regıon can be turned out by the heurıstic search has 1/64 size of the whole area.

A spatial heuristic search module

In artificial intelligence terminology, expert system is a type of software that can solve user's queries based on utilization of knowledges. For implementing a spatial heuristic search in GIS, a simple search module has be build in the GIS The search module is essentially a simple analogue of expert systems.

When the module receives an user's query, it selects corresponding algorithms called 'rules' in terminology of AI. Then, the module executes these rules in the knowledge base The results of the search module are candidate region(s), where the user's query may be satisfied with great possibility

SPATIAL HEURISTIC SEARCH MODULE

The search module, similar to an expert system, includes three components· a control strategy, a set of roles, and a knowledge base. The knowledge base is the spatial knowledge base, mentioned above.

Control strategy

The heuristic search is implemented on the QTSS tree. only one kind of user's queries is dealing with, it is called as "EXTREME" type. User asks that find out the region(s) with the extreme value of a certain geographic measure, such as the highest density region, or the largest connective area, etc Different queries ask for different search methods Most of them are following up-down breath-first strategy(Nilsson, 1983).

Searching algorithms (rules)

For example, two sets of rules are described in more detail

Largest connective block finding --- An up-down breath-first search procedure is executed This is a query of "EXTREME" type, so only one candidate node at each level is enough The first candidate node is the root of the QTSS tree. The evaluation function QAREA (Eqn 2) is implemented on all sons of the candidate node to find the node with the maximum value This node is the new candidate node at this level. Recurring this procedure until arriving a given level L_g. The candidate node at L_g is the final region If a candidate node has nothing, the search procedure stops because there is not any the phenomenon at all in the area. The set of these rules is listed here in a formal representation

(1). The search procedure is implemented in the QUAD tree of the phenomenon.

(2) At the beginning of the search procedure, the root node is the first candidate node.

(3) One candidate node at the level L has four son nodes at level L+1. the new candidate node at level L+1 is the one who has the maximum value of the area function "QAREA(QTSS)".

$$QAREA(QTSS_p(L+1,I))=MAXIMUM(QAREA(QTSS_p(L+1,i))) \quad where\ i=1,2,3,4$$

where the quadrant I is the new candidate node at the level L+1. This rule can be used recursively.

(4). At the given level L_g, the search procedure stops The candidate node at this level is the final candidate region.

(5). If the QTSS of a candidate node have zero value from "QAREA" function, The search procedure stops. The result is that there is not any such phenomenon in the whole area.

(6). At the level L, if two or more than two nodes have same value from "QAREA(QTSS)". They are juxtaposed as candidate node at level L. All their sons of these candidate nodes have to be evaluated for chose the new candidate node at next level. This rule can be used recurse.

Highest density area --- This is also a query of "EXTREME" type. Its search procedure is more simple. From the QTSS tree, all nodes at the given level L_g are directly compared each other. The node who has the maximum value from "QAREA($QTSS_p(L_g,I)$)" is the final candidate node. A formal representation of these rules is listed here.

(1). The search is executed in the QTSS tree of this phenomenon. The candidate node is among the all nodes of the QTSS tree at the given level L_g.

(2). The candidate node has the maximum value from "QAREA".

$$QAREA(QTSS_p(L_g,I)) = MAXIMUM\ (QAREA(QTSS_p(L_g,N))\quad where\ N\ is\ all$$

(3). If the candidate node has zero value from "QAREA", the result is that There is not any such phenomenon at the whole area.

RESULTS ANALYSIS

The test GIS covers an area about half size of LANDSAT images. It is at Black Hill, South Dakota, USA. There are 2048*2048 pixels covering this area with 50*50 meter resolution. The GIS is working at the environment of VAX750/UNIX4.2.

For test time efficiency, the whole test area is cut into 8*8 grids. The user's query is to find the highest density block among 64 blocks. This procedure has only a few of steps for each pixel in a raster image However, it has to calculate 64 areas from 64 block, and compare them to find the largest one. The CPU time spent to calculate the area of each block is about 0 0275 seconds. The whole procedure takes about CPU time 1.760 seconds(TA) As a comparison, the heuristic search to find out the candidate block needs 0 17 seconds(TC). Plus the CPU time to calculate the candidate block, the whole procedure takes 0.195 seconds(TH). The ratio of 0.195/1.760 shows that the improvement for time efficiency is obvious.

Second, the correction of the heuristic search is tested. Heuristic search does not always guarantee a correct result, But it often turns out better results(Pearl, 1984). Thus, besides the CPU time advantage, the correct probability of its results is tested. Among 23 types of land-use in the test area, the heuristic search is used to find the block with a highest density all types have correct results. While the heuristic search is used to find the largest connect block, 22 types have correct results. It means that this heuristic search usually gives right answers but not always.

Figure 2 shows the finding steps of the highest density block on the crops area. Figure 3 shows the finding steps of the largest connective block on the mixed forest area. Those candidate blocks at top several levels have been plotted on both figures. They show the procedure from finding rough approximation candidate blocks to point out more precise candidate blocks.

CONCLUSIONS

This research has provided a fast spatial heuristic search for large scale geographic information systems. It shows the potential feature of utilization of artificial intelligence in GIS. The quad tree spatial spectra(QTSS) is shown its capability of compact way to extract spatial information which can guide spatial

Figure 2. The candidates of the highest density blocks generated by the heuristic search at level 1, 2, and 3 of crops area at Black Hill

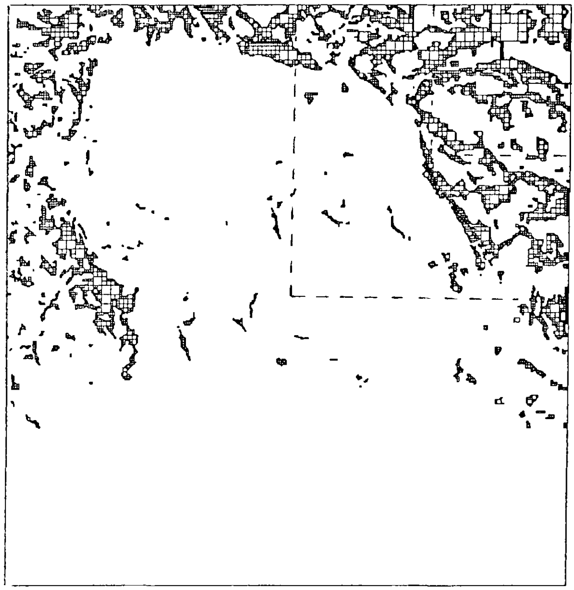
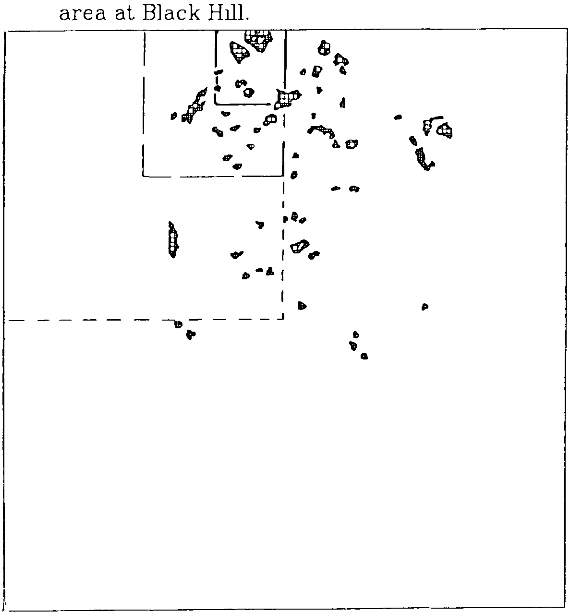


Figure 3. The candidates of the largest connective blocks generated by the heuristic search at level 1, 2, and 3 of mixed forest area at Black Hill.



heuristic search for several kind of queries. The spatial heuristic search avoids the disadvantages of the conventional blind search. It trys finding the higher potential candidate area more directly based the knowledges

From the above test results, the sum of the time spend for finding the candidate block and the time executing on the block is much less than the conventional blind search time. The correction probability is also acceptable

The further research will pay more attention on the algorithms of using spatial knowledges to find different kind of candidate blocks.

REFERENCES

Chen, Z T., 1984, "Quad Tree Spatial Spectra --- Its generation and applications", Proceedings of the International Symposium on spatial data handling, Zurich, Vol 1, pp 208-237.

Klinger, A and Dyer, C D., 1976, "Experiments in picture representation using regular decomposition", Computer Graphics and Image Processing, Vol 5, pp 68-105

Nilsson, N J., 1980, "Principles of Artificial Intelligence", Tioga Publication Co., California.

Pearl, J., 1984, "Heuristics - Intelligent search strategies for computer problem solving", Addison-Wesley Publs. Co , London.

Peuquet, D., 1984, "Data structure for a knowledge-based Geographic Information System", Proceedings of the International Symposium on spatial data handling, Zurich, Vol 2, pp.372-391.

Smith, T R. and Pazner, M , 1984, "Knowledge-based control of search and learning in a large-scale GIS", Proceedings of the International Symposium on spatial data handling, Zurich, Vol.2, pp.498-519.