THE USE OF QUADTREES IN GEOGRAPHIC INFORMATION SYSTEMS AND SPATIAL DATA HANDLING

David M. Mark
Department of Geography
University at Buffalo
Buffalo, New York 14260 U.S.A.

ABSTRACT

Quadtrees represent 2-dimensional (spatial) data in a way which takes advantage of spatial coherence in the phenomena being represented. A square enclosing a region of interest is divided into four quadrants, and these quadrants are recursively partitioned into subquadrants until all subquadrants are uniform with respect to the phenomenon. This paper explores the utility of quadtrees as a data structure for geographic information systems. Attention is focused on linear quadtrees, and in particular on 2-dimensional run-encoding, an effective structure for quadtrees.

INTRODUCTION

A **Geographic Information System** (GIS) is a computerized, spatially-referenced data base organized in such a way that spatial data input, analysis, and output may be accomplished. As is the case for any computer application, the issue of **data structures** is a critical one. Once a data structure for a GIS has been adopted, it is very difficult to change it; data structures also are a major factor in determining the efficiency with which queries can be answered within a GIS.

Recently, **quadtrees** have received considerable attention as a data structure for GIS applications (Mark and Lauzon, 1984; Peuquet, 1984; Samet et al., 1984; Lauzon et al., 1985). Quadtrees appear to have many advantages for handling coherent ('blocky') spatial data, but are inefficient for continuous surfaces such as topography. However, if quadtrees are to be used for natural resources GIS, it is essential to develop strategies for efficient integration of digital elevation model (DEM) data into the quadtree environment; such strategies will be discussed herein.

This paper presents a variety of issues related to a quadtree-based GIS, emphasizing: approaches for handling diverse spatial data types in a quadtree environment; strategies for covering very large areas; and applications of quadtrees and quadtree-related structures to problems in computational geometry, spatial search, and spatial modelling.

Quadtree Definitions and Basic Concepts

A quadtree is a spatial data structure based on a regular
decomposition of an image into quadrants and subquadrants.
This decomposition is conveniently represented by a rooted
tree of out-degree 4 and depth n, where n is the resolution
parameter. The level-0 nodes, or pixels, are conventionally
taken as having a side length of one, and the full region
has a side length of 2^n. Each node of the tree corresponds
to a sub-quadrant of the region. A polygon can be mapped
onto the region by colouring the quadtree for the region
(see Figure 1). By convention, a node is coloured BLACK if
it lies wholly within the polygon, WHITE if wholly outside
the polygon, and GRAY otherwise. Any sub-quadrant of
level > 0 which contains areas of both colours is termed a
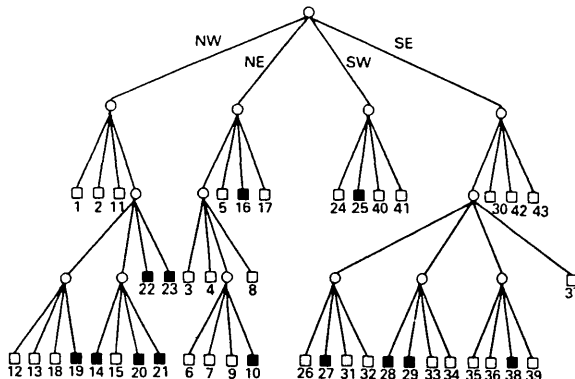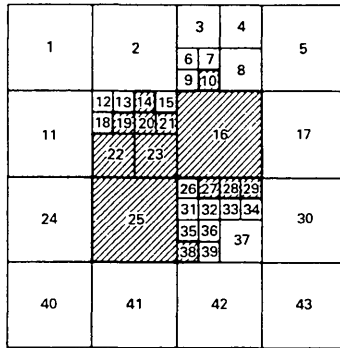GRAY node, and is divided into subquadrants.

Figure 1:   A polygon (A) and its associated
quadtree, represented as a tree (B).

Quadtree Data Structures

Early work on quadtrees represented relations among nodes by an explicit tree structure, with nodes linked to parents and children by pointers (see Samet, 1984, for a review). Another class of data structures, termed **linear quadtrees** (Gargantini, 1982; Mark and Lauzon, 1984; Samet et al., 1984; Mark and Abel, 1985; Samet and Tamminen, 1985) represents only leaf nodes, with those nodes identified by numeric keys. The form of these keys permits topologic and spatial relations to be determined from the key values through bit manimulations or modular arithmetic. The data structure is a list of the leaf nodes, in sequence by key.

Several space-efficient forms of linear quadtrees have been reported in the literature. Gargantini (1982) and Abel (1984) represent explicitly only the BLACK nodes of the tree, inferring WHITE nodes when required. Lauzon and Mark (Lauzon, 1983; Mark and Lauzon, 1984; Lauzon et al., 1985) have proposed a method termed **2-dimensional run-encoding** (2DRE); in this approach, runs of nodes of the same colour which are consecutive in key sequence are represented by the last key within the run. For many algorithms of importance in geographic information systems applications (for example, overlay), 2DRE files can be used directly, without decoding. For other applications, the 2DRE file must be decoded into discrete quadtree leaves.

Mark and Lauzon (1985b) and Lauzon et al. (1985) have shown that 2DRE quadtrees usually require less space and fewer records than do other forms of linear quadtrees, which in turn are more space-efficient than pointer-based quadtrees. Samet (1984) correctly noted that space-efficiency is not always the most important measure of a data structure. However, overlay and certain other GIS algorithms can be solved in a time which is linear in the number of records stored; if a quadtree is represented by fewer records, it often will require less processing time.

ORDERING TWO-DIMENSIONAL SPACE

One of the most important issues in Geographic Information Systems (GIS) is this: geographic data are essentially 2-dimensional, whereas computer storage and processing are (thus far) essentially one-dimensional. No linear sequence can preserve all spatial properties of geographic data; however, the study of such 'orderings' of 2-dimensional space is an important yet neglected topic in GIS research.

An **ordering** of a set of N distinct spatial entities (polygons or points) may be defined as any one-to-one assignment of the integers 1 through N (or 0 through N-1) to the spatial entities. Thus, whereas the 2DRE data structure was originally developed as a compact representation of a

linear quadtree (Lauzon, 1983), 2DRE can also be viewed as an effective ordering of the pixels in a 2-D digital image. In fact, the overlay algorithm presented by Lauzon and Mark (Lauzon, 1983; Mark and Lauzon, 1984; Lauzon et al., 1985) will work for any ordering of an image, including raster (row-by-row), reversing raster (row-prime; Goodchild and Grandfield, 1983), and Hilbert-Peano (or pi; Goodchild and Grandfield, 1983) ordering, as well as the Morton (or N; White, 1983) ordering used in 2DRE.

One of the strengths of the 2DRE representation using the Morton sequence is its dual nature: it can be treated either as a 2-D ordering or as a compact linear quadtree. A disadvantage of the Morton sequence it that it does not traverse the image in a spatially-contiguous path. The Hilbert-Peano (pi) order may prove to be valuable for GIS applications, since, like row-prime, it always moves to a 4-neighbour, and yet, like Morton, pi-order visits all pixels in a quadtree subquadrant before leaving it. Goodchild and Grandfield (1983) found that images could be represented in fewer runs using pi-ordering than using Morton ordering. However, the recursion implicit in a quadtree is straight-forward for Morton order and complicated for pi-order. The use of pi-order in GIS has yet to be explored.

DATA TYPES AND THEIR REPRESENTATIONS IN A QUADTREE-BASED GIS

It is proposed to recognize four fundamental classes of data, based mainly on dimensionality. For each quadtree area, several files may exist, each containing a different data type.

Type 0: Point Files
Point files would be held in Morton number. Point locations would be reported to the nearest pixel or grid cell and converted to within-patch pixel coordinates. The coordinates would then be combined to form the Morton number (key) of the point. Each record in the key-ordered file would contain the key of the point, its attributes and/or values, and perhaps its coordinates. Points would be held in key-order for ease of interfacing with other data types, and also because White (1983) and Abel and Smith (1984b) have shown that this is an effective data structure for solving closest point problems and other problems in computational geometry.

Type 1: Line Files
It is well-known that quadtrees do not handle line data well. However, a modified quadtree structure termed the **nonminimal division quadtree** has recently been proposed by Ayala et al. (1985); this may be the best method for

handling geographic line **features** such as power lines, roads, and streams. Note however that **boundary** lines would not be stored as line data; rather, the bounded regions are stored as coverage files (Type 2, below).

Type 2: Coverage Files
Coverage data, such as land use, soils, or rock types, will be stored as linear quadtree files. In most if not all cases, the 2-dimensional run-encoding method for file structuring is recommended for these space-filling coverages; the value of this structure already has been discussed above.

Type 3: Surface Files
Geographic information systems frequently involve natural resource data. In most such applications, elevation data (digital elevation models) represent an important component, both directly and in the form of derived measures such as slope. Thus, any data structure adopted for a resources GIS must be able to efficiently integrate DEM data with other geographic data.

Quadtree representations are not efficient if neighbouring cells seldom have identical values. This is, however, a characteristic of most DEMs, and of LANDSAT or other MSS satellite imagery. For such data, it is more efficient to store a value for every grid cell. For ease of interfacing with other data types, the proposed system would use the Morton number (key) of each pixel as a virtual address for referencing the elevation within a contiguous binary file (Lauzon et al., 1985; Cebrian et al., 1985). Cebrian et al. (1985) presented details of a strategy for integrating such data into a quadtree-based GIS, and for performing basic DEM analysis and display procedures.

Relating Type 2 (Thematic) to Type 3 (DEM) data
As noted above, the combination of Morton-order DEM files and 2DRE quadtree files provides the basis for efficient interfacing of the two types of data. Two different display procedures for illustrating associations between topography and other data have been devised and implemented (Cebrian et al., 1985). The first displays topographic properties (heights, slopes, slope aspects, etc.) within only those areas having some particular property or properties (such as a certain land cover or soil type) in one or more coverage files. The search area is defined by one or more 2DRE files. The second overlay procedure produces graphic output by overlaying thematic data on an image produced by analytical hill shading. In order to produce such a display, a special palette must be defined and loaded into the graphics device. This palette consists of several series of colours; each

series has a particular hue, but with different **values** (see Cebrian et al., 1985).

HANDLING VERY LARGE AREAS IN A QUADTREE-BASED GIS

Most real-world applications of GIS involve large areas. If a single coordinate system is used to cover such a very large area, either a large number of bits must be used for each coordinate, or precision must be lost. To avoid both of these undesireable effects, a GIS for a large area can partition the space into a set of mutually-exclusive and collectively-exhaustive **patches** (also known as frames). Such a system can be accommodated very easily within a quadtree-based GIS, since the same recursive system of quadrants can be used both within the patches and above them. Quadtrees for large areas do encounter problems when applied to sufficiently large portions of the Earth's surface, since squares cannot tesselate a sphere; thus, application of quadtrees to global or continental regions requires either the explicit use of map projections, or the modification of the quadtree concept to use quadrants which are not geometrically square.

Quadrangle-based Systems
One class of solutions to the problem involves partitioning the globe into areas which may be termed **quadrangles**, cells which are 'square' or 'rectangular' only in latitude-longitude terms. Such systems have a variety of analysis and display problems, since cells are not geometrically square, and furthermore change size and shape with latitude. However, the frequent use of latitude-longitude quadrangles for non-computerized mapping makes them attractive, since printed maps often are used as a primary data source for GIS.

A good example of a quadrangle-based GIS which is hierarchical in the quadtree sense (but only down to a certain scale) is the **Canada Geographic Information System** (Tomlinson et al., 1976). CGIS was the first full geographic information system. Indeed, although the quadtree idea is usually attributed to Klinger's 1971 paper, many quadtree-related concepts were included in CGIS in the mid-1960's (Morton, 1966; Tomlinson et al., 1976).

Systems Based on Map Projections
Another class of solutions maps the globe onto a plane or set of planes, using some map projection, and then defines a grid cell network in cartesian coordinates on the plane(s). Since no map projection can be both equal-area and conformal, square cells on the map would represent areas in the real world which vary in size, shape, or both. Mark and

Lauzon (1985a) proposed that a continental or world-wide scale GIS based on quadtree concepts should use the Universal Transverse Mercator (UTM) coordinate system. In such a system, three hierarchical levels would be used. The highest divides the world into UTM zones and subzones. Each UTM subzone is then divided into square patches, which are numbered according to the Morton. Finally, within each patch, a 256 by 256 array of cells is the basis for the quadtrees or other geographic data files.

For each patch, a variety of data sets may exist, using any or all of the four fundamental data types discussed above. The highest level of the GIS would consist of a data base management system (DBMS) which would contain a directory of patches, data types, and data sets actually available, with summary statistics relating to the file contents. In fact, the patches themselves can be treated as pixels, and summary statistics can be mapped at this highly generalized level.

The use of the UTM coordinate system is recommended for a quadtree-based GIS because: conversion of geographic (latitude-longitude) coordinates to UTM is well known and computer programs or formulae are readily available; UTM coordinates are in general use by the armed forces of the United States, Canada, the United Kingdom, and other countries; and the U.S. Geological Survey (USGS) distributes digital data either already in UTM coordinates or with coefficient for conversion to UTM contained in the file headers (see Elassal and Caruso, 1983; Alldder and Elassal, 1984). For further details of the use of UTM coordinates in quadtree construction, see Mark and Lauzon (1985a).

QUADTREES, COMPUTATIONAL GEOMETRY, AND SPATIAL MODELLING

Computational geometry seeks efficient solutions to geometric problems. Many problems in spatial data handling stem from the fact that geographic space is 2-dimensional, whereas most computer processing is 1-dimensional (see discussion above). As an example, consider the problem of finding a point's nearest neighbour. To the human eye, candidates for the 'nearest neighbour' are obvious, and only a few measurements need be made to ascertain which point is indeed the closest. Limiting the amount of unnecessary searching is a central theme in computer handling of spatial data.

Quadtrees and related spatial data structures have considerable potential in this regard. Indeed, the region quadtree was first suggested (by Klinger, 1971) in the context of spatial search, and not as a data structure for images. Samet (1984, pp. 229-244) has provided a detailed review of methods for handling point data and rectangle data using quadtree-related methods. Abel and Smith (1983;

1984a) have discussed quadtree-based solutions to the rectangle retrieval and rectangle cover problems, which arise in applications as different as VLSI architecture and GIS. They also showed (Abel and Smith, 1984b) how linear quadtrees can be used for nearest neighbour calculations.

Quadtrees also hold particular promise for spatial modelling. For example, Mark (in prep.) shows how calculation of proximal (Thiessen) polygons can be simplified through the use of quadtrees. The basic algorithm is recursive: if all four corners of a square have a common closest point, then that square is part of the proximal polygon surrounding that point, and is part of the quadtree of the proximal polygon map. Otherwise, the current square is split into four sub-quadrants, and the algorithm tests each of these. The algorithm works because proximal polygons are convex. A recursive, quadtree-based approach may prove valuable in other spatial modelling situations.

SUMMARY

Quadtrees are very well-suited to many Geographic Information Systems (GIS) applications, chiefly because they represent 2-dimensional (spatial) data in a way which takes advantage of spatial coherence in the phenomenon being represented. This paper has emphasized the handling of diverse types of spatial data in a quadtree environment, strategies for covering very large areas, and the use of quadtrees and quadtree-related structures in computational geometry, spatial search, and spatial modelling.

ACKNOWLEDGEMENTS

REFERENCES

Abel, D. J., 1984, A B^+-tree structure for large quadtrees. Computer Vision, Graphics, and Image Processing, v. 27, pp. 19-31.

Abel, D. J., and Smith, J. L., 1983, A data structure and algorithm based on a linear key for a rectangle retrieval problem. Computer Vision, Graphics, and Image Processing, v. 24, pp. 1-13.

Abel, D. J., and Smith, J. L., 1984a, A data structure and query algorithm for a database of areal entities. Australian Computer Journal, v. 16, pp. 147-154.

Abel, D. J., and Smith, J. L., 1984b, A simple approach to the nearest-neighbour problem. Australian Computer Journal, v. 16, pp. 140-146.

Allder, W. R., and Elassal, A. A., 1984, Digital line graphs from 1:24,000-scale maps. U. S. Geological Survey, Circular 895-C, 79pp.

Ayala, D., Brunet, P., Juan, R., and Navazo, I., 1985, Object representation by means of nonminimal division quadtrees and octrees. ACM Transactions on Graphics, v. 4, pp. 41-59.

Cebrian, J. A., Mower, J. E., and Mark, D. M., 1985, Analysis and display of digital elevation models within a quadtree-based geographic information system. Proceedings, Auto-Carto 7, pp. 55-65.

Elassal, A. A., and Caruso, V. M., 1983, Digital elevation models. U. S. Geological Survey, Circular 895-B, 40pp.

Gargantini, I., 1982, An effective way to represent quad-trees. Communications of the ACM, v. 25, pp. 905-10.

Goodchild, M. F., and Grandfield, A. W., 1983, Optimizing raster storage: an examination of four alternatives. Proceedings, Auto-Carto 6, v. 1, pp. 400-407.

Klinger, A., 1971. Patterns and search statistics. In Rustagi, J. J., editor, Optimizing Methods in Statistics. New York: Academic Press.

Lauzon, J. P., 1983, Two-dimensional Run-encoding for Spatially Referenced Data. Unpublished M.A. project, Department of Geography, State University of New York at Buffalo.

Lauzon, J. P., Mark, D. M., Kikuchi, L., and Guevara, J. A., 1985, Two-dimensional run-encoding for quadtree represent-ation. Computer Vision, Graphics, and Image Processing, v. 30, pp. 56-69.

Mark, D. M., Recursive algorithm for determination of proximal (Thiessen) polygons in any metric space. Manuscript in preparation.

Mark, D. M., and Abel, D. J., 1985, Linear quadtrees from vector representations of polygons. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7, pp. 344-349.

Mark, D. M., and Lauzon, J. P., 1984, Linear quadtrees for geographic information systems. Proceedings, International Symposium on Spatial Data Handling, Zurich, Switzerland, August 1984, vol. 2, pp. 412-430.

Mark, D. M., and Lauzon, J. P., 1985a, Approaches for quadtree-based geographic information systems at continental or global scales. Proceedings, Auto-Carto 7, pp. 355-364.

Mark, D. M., and Lauzon, J. P., 1985b, The space efficiency of quadtrees: an empirical examination including the effects of two-dimensional run-encoding. Geo-Processing, v. 2, pp. 367-383.

Morton, G., 1966. A computer oriented geodetic data base, and a new technique in file sequencing. IBM Canada Limited, unpublished report, March 1, 1966.

Peuquet, D. J., 1984, Data structures for a knowledge-based geographic information system. Proceedings, International Symposium on Spatial Data Handling, Zurich, Switzerland, August 1984, vol. 2, pp. 372-391.

Samet, H., 1984, The quadtree and related hierarchical data structures. Computing Surveys, v. 16, pp. 187-260.

Samet, H., Rosenfeld, A., Shaffer, C. A., and Webber, R. E., 1984, Use of hierarchical data structures in geographic information systems. Proceedings, International Symposium on Spatial Data Handling, Zurich, Switzerland, August 1984, vol. 2, pp. 392-411.

Samet, H., and Tamminen, M., 1985, Computing geometric properties of images represented by linear quadtrees. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7, pp. 229-239.

Tomlinson, R. F., Calkins, H. W. and Marble, D. F., 1976, Computer Handling of Geographical Data: An Examination of Selected Geographic Information Systems. Paris: The UNESCO Press.

White, M., 1983, N-trees: Large ordered indexes for multi-dimensional space. Paper presented, Lincoln Institute of Land Policy's Colloquium on Spatial Mathematical Algorithms for Microcomputer Land Data Systems, May 9-10, 1983, Boston, Massachusets.

LANDSAT AND SPOT HIGH RESOLUTION SATELLITE IMAGES :

A NEW COMPONENT FOR GEOGRAPHIC DATA BASES

J. Denègre

J.C. Deschamps

B. Galtier

Institut Géographique National - France

136 bis rue de Grenelle

75700 Paris

ABSTRACT

Satellite imagery with infra-hectometric ground resolution constitutes an apport without former example for automation in cartography. Satellites represent first of all an automatic way for data acquisition, due to their mondial extension capacity and to their repetitivity, largely improved by the possibility of oblique views (SPOT) and stereoscopy. But they bring also a very important advantage for processing and representing geographical data.

After centuries during which the draftman's hand, relayed by printing techniques, attempted to represent the real world on maps made with lines and points, satellite imagery supported by computer facilities brings suddenly a two-dimensional representation, made with surfaces and directly usable as cartographic material.

It may be of interest to think about the primary function of topographic base maps : representing all features existing on the Earth surface, and the form of it (altimetry), they constitute basically a framework through which everyone can recognize and orientate himself in his environment. From that point of view, and particularly in new countries with little planimetric density, where natural features are the majority (and how difficult to compile with conventional techniques !), satellite images after rectification and stereo-compilation for altimetric survey can play the role of a fundamental geographic background.

Together with the possibility of semi-automatic interpretation for up-dating certain topics of conventional maps and for the inventory of natural resources, the availability of new mass storage media like the digital optical disk makes it possible for satellite images to take place into geographic data bases.