International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.4., 2015
(July-Aug)

RESEARCH ARTICLE

ISSN: 2321-7758

IMPROVING GIS SPATIAL QUERY PROCESSING ON SPATIAL DATABASE USING HILBERT R-TREE AND B-TREE

NEERAJ KUMAR DUBEY[1], Dr. SANJAY AGRAWAL[2]

[1]Department of Computer Technology and Application, M.Tech Scholar NITTTR, Bhopal., M.P India

[2]Department of Computer Technology and Application, Professor and Dean of Research NITTTR, Bhopal., M.P, India

NEERAJ KUMAR DUBEY

ABSTRACT

The size of spatial database is inversely proportional to the efficiency of information retrieval. The potency of spatial range query in geographic information systems (GIS) is of predominant importance and it depends on the query processing algorithms. When the database sizes increases, the processing of data retrieval from spatial database gets complicated. This paper merges two type of data structures; Hilbert R-tree and Balanced tree (B-tree) called as Hilbert RB-tree. In this paper Hilbert RB-tree utilizes the efficiency of Hilbert R-tree and B-tree collectively and accelerates execution speed of complex queries and minimizes the time consumed in searching, editing, deleting, updating any record in the spatial database. In this paper we propose a framework to test the efficiency of spatial query processing using both data structure.

Key Words— spatial database, spatial indexing, Hilbert R-tree, B-tree, Hilbert RB-tree.

INTRODUCTION

Geographic information system is used to locate and access geospatial data from geospatial database quickly and efficiently. Geographic database stores both spatial and non-spatial data. Spatial indexing can be used to provide fast response to spatial queries. Spatial database contains collection of records which represents spatial objects; each record contains a identifier which is used for retrieving spatial data.

Spatial data objects may contain one or more than one polygon which has been distributed randomly through space. It is complex and infeasible to store groups of spatial objects in one relational table with static record size. Database systems need to indexing technique to retrieve spatial data quickly and efficiently. There are number of indexing technique that had been used for solving multidimensional data problem like-R-tree, R*-tree, R+-tree, Hilbert R-tree etc.

Hilbert R-tree is a advancement of R-tree and uses space filling technique. Hilbert R-tree is more efficient for indexing

of multidimensional data as compare with R-tree and other variants of R-tree. Basically, Hilbert R-tree is used for spatial attributes. On the other hand

B-tree is used for non-spatial attributes. Spatial data known as geometric data or shape data stored in spatial database.

Attribute data are used to describe the features and characteristic of spatial objects. Spatial data are mainly categorized in two formats: raster data and vector. This type of data periodically generated via specialized sensors, satellite or art craft-mounted cameras (sampling geographical regions into digital image).

Geospatial data divided into two groups; the first group is known as documentary group contains documentary data that describes data: it is the combination between numbers, letters, video clips, and images and it describes the whole map. The second group is known as geometric group which contains geometric data and describes location with coordinates (x, y); it represents numerical data. Making complete geospatial database requires two important files: documentary file and geographic file. Geospatial data is defined as geographically referenced data that describes both location and characteristics of spatial data such as roads, land parcel, vegetation stands on the earth's surface. It is called as geospatial data and stored in spatial databases; multidimensional trees are used to build the indices of spatial data. Attributes of spatial objects are one dimensional and it can be stored in relational database with references to the spatial data.

The main purpose of this paper is to enhance the performance and processing speed of spatial queries. This will be done by merging two data structure namely Hilbert R-tree and B-tree. The proposed system integrates the Hilbert R-tree and B-tree models to facilitate executing complex queries and minimize the time consumed in searching, editing, deleting, updating any record in spatial database compared with the B-tree or Hilbert R-tree alone. In the aggregate Hilbert RB-tree the extents of all regions are stored in Hilbert R-tree. Each leaf and non-leaf entry of Hilbert R-tree is associated with a pointer to B-tree that stores historical aggregate data about the entry.

The rest of the paper is organized as follows: section II discusses the background and related work; section III describes the proposed framework; and finally last section IV contains the conclusion.

Background and related work

GIS is a system of software and hardware which is used for storage retrieval, mapping and analysis of geographic data. Due to large size of geo database and the long processing steps to retrieve these database require a technique for faster and easier retrieval of spatial database. Indexing technique can be used for this purpose; indices will increase speed of search in spatial database.

Spatial indexing technique can be understand as the kind of data structure which has been arranged in some way for the position and shape of space entities or spatial relationship between space entities[1]. Another definition of spatial index is "the primary key used to improve the efficiency of spatial query [9]".

Hilbert R-tree is a dynamic index structure for spatial database. Hilbert R-tree is a advancement of R-tree which uses Hilbert curve or space filling curve.

We use space filling curve for reducing the dimensionality of multiple dimensional data. It is used to maps data from high dimensional space to one dimensional space [3,4]. Figure 1 shows Hilbert space filling curve. We construct Hilbert R-tree on the basis of Hilbert value.

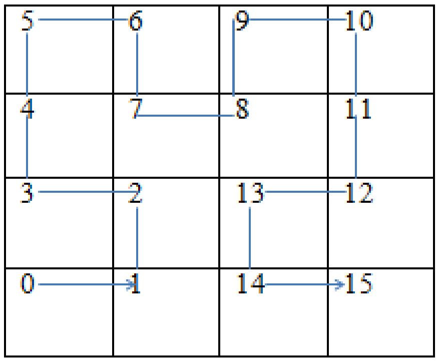

Figure: 1 Hilbert curve

Hilbert R-tree Overview

Based on the previous research, there are number of spatial data indexing technique proposed. Hilbert R-tree is the most widely accessed method used. Zhang et.al in [5] proposed a novel algorithm that answers the kNG query by using Hilbert R-tree. They proved that the Hilbert R-tree is better in performance as compare with R-tree.

To improve the efficiency of query, we use Hilbert R-tree. There are two type of Hilbert R-tree one for static databases and another for dynamic databases. The performance of any R –tree is depends on the used algorithm for clustering the data rectangles on a node. Hilbert R-tree uses space filling curves. The Hilbert curves are used to achieve better ordering of multidimensional data in the node. This ordering is good in the sense that it bind similar data rectangle together to minimize the area and perimeter of final minimum bounding rectangle (MBRs).

The Hilbert R-tree can achieve a high degree of space utilization as required. Hilbert value (the Hilbert value of a point is the length of the Hilbert curve from the origin to the point) of the center of rectangle (MBRs) is used to sort the rectangles. By the use of ordering, every node has a well-suited sibling nodes; thus deferred splitting can be used.

Hilbert R-tree structure

In The Hilbert R-tree a leaf node contains the entries in the form of (R, OBJ_ID) and it may contain at most C1 entries where C1 is the capacity and R is the minimum bounding rectangle (MBR) of real object having co-ordinate (X_L,X_H, Y_L, Y_{HI}) and OBJ_ID is a pointer which points to the object description record.

Non-leaf node in Hilbert R-tree maintains the entries in the form (R, ptr, LHV) and contains at most C_n entries where C_n shows the capacity of non-leaf node, R is the MBR that maintains all the children of that node, ptr shows the pointer to child node, and LHV shows the largest Hilbert value among the data rectangles enclosed by R.

Figure 2 shows a Hilbert R-tree which is built on a data sets R. The total space is partitioned into 16 cells with the help of Hilbert space filling curve. The structure of Hilbert R-tree is nearly same as R-tree. In the Hilbert R-tree all the points are ordered on the basis of Hilbert value. Figure 2 shows the fundamental structure of Hilbert R-tree in which whole space is divided into four regions and each region contain some points or non-spatial attributes.

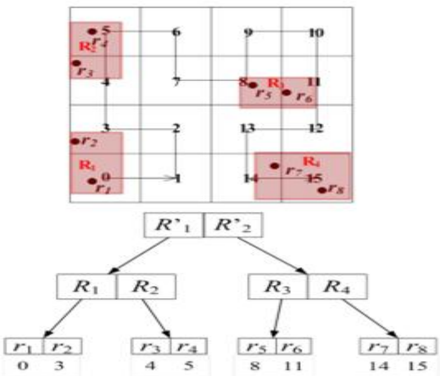

Figure: 2 an example of Hilbert R-tree [12]

B-tree Overview

B-tree is a dynamic data structure which is used to store and retrieve data. B-trees are also called multi-ways trees in which each node contains keys and pointers. General structure of B-tree is shown in Figure 3. B-tree is made of nodes where each node contains left pointer, right pointer and key. The root pointer points the first node of B-tree and left pointer points the left sub-tree and right pointer points the right sub-tree.

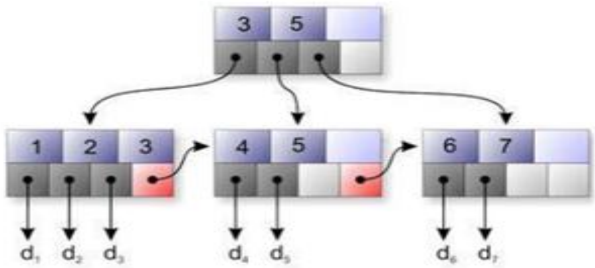

Figure: 3 B-tree Structure [10]

The height of B-tree is always order of log n that is O (log n). B-tree is a dynamic data structure so the height of B-tree grows and shrinks as records are added or deleted from B-tree.

Hilbert RB-tree

In the Hilbert RB-tree, the extents of all regions which are r_1, r_2, r_3, r_4 ...r_8 are stored in Hilbert R-tree. Each leaf and non-leaf entry of Hilbert R-tree is associated with a pointer to a B-tree that which stores aggregate data about the entry. Figure 4: shows the general structure of Hilbert RB-tree.

Different researches have already used RB-tree as an indexing technique in geo-database. In research [11] combination of R-tree and B-tree are used for approximate query processing to solve the problem of distinct count. In research [1] combination of R-tree and B-tree is used to enhance the spatial queries processing.

In this paper we are using advanced or improved R-tree which is known as Hilbert R-tree. Hilbert R-tree uses space filling curve to utilize the space and reduces the dimensionality of multiple dimensional data. We are using combination of Hilbert R-tree and B-tree to reduce the space and time complexity of spatial queries processing.

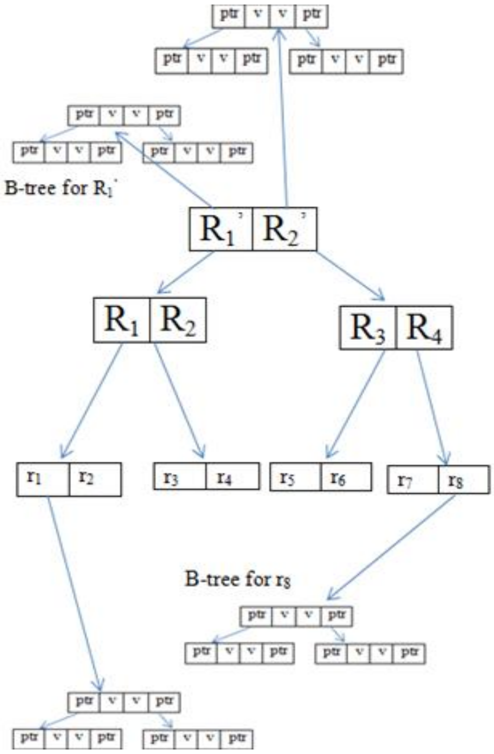

Figure: 4 structure of Hilbert RB-tree

PROPOSED FRAMEWORK

This section mainly discusses the proposed framework that contains the Hilbert RB-tree module, the reason to use Hilbert R-tree and B-tree, and the System design and analysis phase details.

The proposed model uses two data structure; first is Hilbert R-tree implementation and second is B-tree implementation. These two techniques are merged to provide efficient way of searching and provide him full and complete information. The user communicates with web application interface and enters the required address and defines the range of the miles to search within.

Reason to use Hilbert R-tree and B-tree

The goal of integration and use of Hilbert R-tree and B-tree is to provide efficient storage and search facility to user and to provide full information to the user in terms of location and addresses he is looking for in efficient way.

In the proposed framework we are using two modules. In the first module user will interact with web interface for searching some addresses and some non-spatial attributes. The first module starts working by searching the required address in Google map. After that second module will be start. Then Second module searches the non-spatial attributes in system database. Figure 5 shows the working of proposed framework.

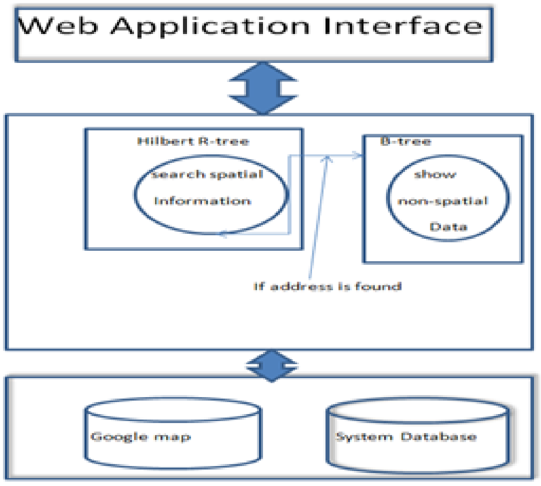

Figure: 5 System Design

Analysis and Design

The main function and activity of proposed framework are described by use case diagram in figure: 6. In figure:6 seven use cases are used which are: search address, compare address, set search query, get address, plot address point, search address and search address details as shown in use case diagram.

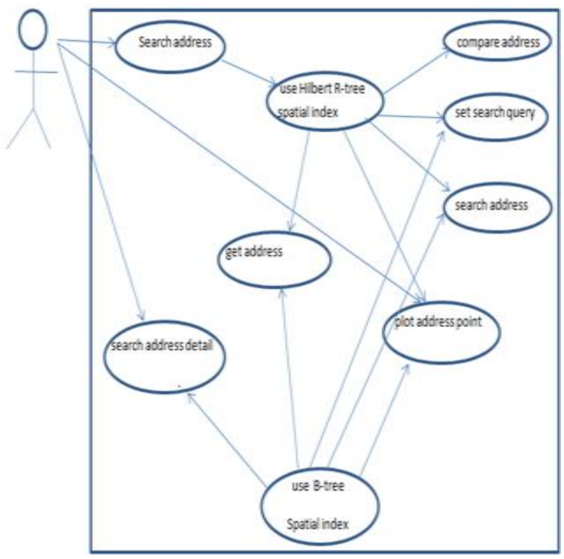

Figure: 6 use case design of system

CONCLUSION

For retrieving and manipulating geospatial data efficiently and effectively, it should be taken into consideration that each system requires a special indexing technique by which geo spatial data can be retrieve quickly. The traditional indexing techniques for geospatial databases are unsuitable and inefficient. As previously shown, Hilbert RB-tree indexing method meets these requirements and provide an efficient and optimized way for accessing multidimensional geospatial data. The important functionality of Hilbert RB-tree is to be capable to retrieve any data quickly and effectively according to its spatial location. In this paper we used to well-known data structure and Hilbert R-tree and B-tree and propose a web based frame work to test the efficiency of Hilbert RB-tree.

REFERENCES

[1] Z. Wei, W. Wanzhen, Y. Xingguang, X. Gang, "An Optimized Query Index Method Based on R-tree," in Proc Fourth International Joint Conference on Computational Sciences and Optimization, pages:1007-1010, (2011).

[2] Y. Lifang; L. Rui; H. Xianglin; L. Yueping, "Performance of R-tree with slim-down & Reinsertion Algorithm," in Proc International Conference on Signal Acquisition and Processing,pages:291-294, (2010).

[3] Z. Shaohui, C. Zhanwei , "The research of Hilbert R-tree spatial index algorithm based on Hybrid clustering," in Proc International Conference on Electronic & Mechanical Engineering and Information Technology,pages:3495-3497,(2011).

[4] Guttman, "R-trees: A Dynamic Index Structure for Spatial Searching", in Proc ACM SIGMOD International Conference on Management of Data, Boston, Ma, pages: 45-57, (1984).

[5] G. Jun K. Shengnan, "A new node-split algorithm in R-tree based on spatial clustering," IEEE Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010) Vol.5, pages: 2081- 2085, (2010).

[6] (2010) The ESRI website.[Online].Available: http://www.esri.com/

[7] J. GONG, Sh. KE,"3D spatial query implementation method based on R-tree," in Proc Remote Sensing, Environment and Transportation Engineering (RSETE), pages: 2828- 2831, (2011).

[8] L. Weihua, W. Yonggang, T. Xiaojun, and Y. Yan "An improvement of index method & structure based on R-tree," IEEE Proc International Conference on Computer Science and Software Engineering, Vol.4, pages: 607- 610, (2008).

[9] Belbaraka, M."Algorithms for generating and coding B-Trees", ProQuest Dissertations and Theses (PQDT), (1996).

[10] Y. Tao, G. Kollios, J. Considine, F. Li, and D. Papadias "SpatioTemporal Aggregation Using Sketches," in ICDE, (2004).

[11] DongEun Kim et.al. "Distributed Processing of Range Queries with Non-Spatial Selections" ISBN: 978-0-9891305-4-7 in SDIWC (2014)

[12] Kao, B. Lee, S. Lee, F. Cheung, D." Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index,"IEEE Transactions on Knowledge and data engineering, Vol.22 No.9, 2010.