Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Information Systems 30 (2005) 205–226
www.elsevier.com/locate/infosys

Neighbor-finding based on space-filling curves ☆

Hue-Ling Chen, Ye-In Chang*

Department of Computer Science and Engineering, National Sun Yat-Sen University, Lien Hai Road, 70, Kaohsiung 804, Taiwan, ROC

Received 24 July 2002; received in revised form 14 November 2003; accepted 22 December 2003

Abstract

Nearest-neighbor-finding is one of the most important spatial operations in the field of spatial data structures concerned with proximity. Because the goal of the space-filling curves is to preserve the spatial proximity, the nearest neighbor queries can be handled by these space-filling curves. When data are ordered by the Peano curve, we can directly compute the sequence numbers of the neighboring blocks next to the query block in eight directions in the 2D-space based on its bit shuffling property. But when data are ordered by the RBG curve or the Hilbert curve, neighbor-finding is complex. However, we observe that there is some relationship between the RBG curve and the Peano curve, as with the Hilbert curve. Therefore, in this paper, we first show the strategy based on the Peano curve for the nearest-neighbor query. Next, we present the rules for transformation between the Peano curve and the other two curves, including the RBG curve and the Hilbert curve, such that we can also efficiently find the nearest neighbor by the strategies based on these two curves. From our simulation, we show that the strategy based on the Hilbert curve requires the least total time (the CPU-time and the I/O time) to process the nearest-neighbor query among our three strategies, since it can provide the good clustering property.

Keywords: Hilbert curve; Neighbor-finding; Peano curve; RBG curve; Space-filling curves; Spatial proximity

1. Introduction

With the proliferation of wireless communications and the rapid advances in geographic information systems (GIS), a very common spatial query is the "Nearest Neighbor Query" which seeks for the objects residing more closely to a given object. For example, when driving a car in a highway, the driver may ask where the nearest gas station is. In other words, nearest-neighbor-finding is one of the most important spatial operations in the field of spatial data structures regarding proximity.

Basically, spatial data consist of objects in space made up of points, lines, regions, rectangles, and data of higher dimensions. Access methods are required to support efficient manipulation of the multi-dimensional spatial objects in the secondary storage. Several schemes [1–7] for the linear mapping of a multi-dimensional space have been proposed for various applications, such as access methods for traditional databases, GIS and spatio-temporal databases. However, in spatial applications, the structure of databases become larger and more complex and means for extracting valuable

☆ Recommended by Nick Koudas.

*Corresponding author. Tel.: +886-7-525-2000x4334; fax: +886-7-525-4301.

E-mail address: changyi@cse.nsysu.edu.tw (Y.-I. Chang).

0306-4379/$ - see front matter
doi:10.1016/j.is.2003.12.002

information become more sophisticated [4]. The design of multi-dimensional access methods is difficult compared to one-dimensional cases because there is no total ordering that preserves spatial locality [5]. Once a total ordering is found for a given spatial database, one can use any one-dimensional access method which may yield good performance for multi-dimensional queries. Thus, what is needed is a mapping from a higher dimensional space to a one-dimensional (1D) space. In this mapping, the elements which are close in space are mapped into nearby points on the line to preserve the spatial locality. A space-filling curve passes through every point in the space once. It is one kind of mapping to give an one-to-one correspondence between the coordinates and the sequence numbers of the points on the curve [4].

Because the goal of the space-filling curves is to preserve spatial proximity, they can handle nearest-neighbor queries. Some examples of space-filling curves are the Peano curve, the RBG curve and the Hilbert curve. The *Peano curve* [6,7] (or the *Z*-order curve) has the *bit shuffling* property, which means interleaving the bits from two coordinates of the point in base 2 in the two-dimensional (2D)-space. An improvement of the Peano curve is the *Reflected Binary Gray-code* (*RBG*) *curve* [1,2] which uses Gray coding on the interleaved bits to reduce random accesses on the disk for range queries. Based on the Peano curve, the *Hilbert curve* was proposed with superior *data clustering* properties, which means that the locality between objects in the multi-dimensional space is preserved in the linear space [3–5,8,9].

When the data are linearly ordered by the Peano curve, we can directly compute the sequence numbers of the neighboring blocks next to the query block in eight directions based on its bit shuffling property, even the *order* is greater than 1. (Note that the order of a curve deals with how many sequence numbers be necessary to uniquely define that curve.) But with the data ordered by the RBG curve or the Hilbert curve, neighbor-finding would be complex. However, we also observe that there is some relationship between the RBG curve and the Peano curve, as with the Hilbert curve. Therefore, in this paper, we first show how these space-filling curves are derived and discuss the relationships between them. Next, we describe the effects on nearest-neighbor queries by using different curves. Then, we show the way to find the nearest neighbor by our strategy based on the Peano curve, and present the rules for transformation between the Peano curve and others. Finally, we compare the performance of our strategy for the nearest-neighbor-finding based on different space-filling curves.

The rest of paper is organized as follows. In Section 2, we briefly describe three space-filling curves, the Peano curve, the RBG curve, the Hilbert curve, and their spatial properties. In Section 3, we present our strategies for the nearest-neighbor-finding, based on the Peano curve and the transformation rules between the Peano curve and the others. In Section 4, we describe the simulation experiment performed to compare different strategies, and results obtained therefrom. Finally, we provide conclusions for this paper.

2. Space-filling curves

The design of multi-dimensional access methods is difficult compared to one-dimensional cases because there is no total ordering that preserves spatial locality [5]. One way is to look for a mapping which preserves spatial proximity at least to some extent. In this section, we discuss the properties of the Peano curve, the RBG curve, and the Hilbert curve.

2.1. Properties

A space-filling curve orders points linearly to preserve the distance between two points in the 2D-space. This means that points which are close in space and represent similar data should be stored together in the linear sequence. Some examples of space-filling curves are the Peano curve, the RBG curve and the Hilbert curve [1–3,5–9]. In general, a space-filling curve starts with a basic path on a *k*-dimensional square grid of side 2. The path visits every point in the grid exactly once without crossing itself. It has two free ends