Information Sciences 180 (2010) 1292–1312

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Efficient processing of spatial joins with DOT-based indexing

Hyun Back [a], Jung-Im Won [b,*], Jee-Hee Yoon [a], Sanghyun Park [c], Sang-Wook Kim [d]

[a] *Division of Information Engineering and Telecommunications, Hallym University, Republic of Korea*
[b] *College of Information and Communications, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, Republic of Korea*
[c] *Department of Computer Science, Yonsei University, Republic of Korea*
[d] *College of Information and Communications, Hanyang University, Republic of Korea*

ARTICLE INFO

Article history:
Received 6 August 2008
Received in revised form 4 August 2009
Accepted 6 November 2009

Keywords:
Spatial databases
Spatial indexing
DOT index
Spatial join
Space-filling curve

ABSTRACT

A *spatial join* is a query that searches for a set of object pairs satisfying a given spatial relationship from a database. It is one of the most costly queries, and thus requires an efficient processing algorithm that fully exploits the features of the underlying spatial indexes. In our earlier work, we devised a fairly effective algorithm for processing spatial joins with double transformation (DOT) indexing, which is one of several spatial indexing schemes. However, the algorithm is restricted to only the one-dimensional cases. In this paper, we extend the algorithm for the two-dimensional cases, which are general in Geographic Information Systems (GIS) applications. We first extend DOT to two-dimensional original space. Next, we propose an efficient algorithm for processing range queries using extended DOT. This algorithm employs the *quarter division technique* and the *tri-quarter division technique* devised by analyzing the regularity of the space-filling curve used in DOT. This greatly reduces the number of space transformation operations. We then propose a novel spatial join algorithm based on this range query processing algorithm. In processing a spatial join, we determine the access order of disk pages so that we can minimize the number of disk accesses. We show the superiority of the proposed method by extensive experiments using data sets of various distributions and sizes. The experimental results reveal that the proposed method improves the performance of spatial join processing up to three times in comparison with the widely-used R-tree-based spatial join method.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Spatial objects are objects that have locations and sizes within space [15]. *Spatial database systems* provide functionalities for the storage and management of a large number of spatial objects and support applications such as Geographic Information Systems (GIS), Very Large Scale Integration (VLSI), and Computer Aided Design/Computer Aided Manufacturing (CAD/CAM) [15,9]. A variety of spatial queries such as *spatial point queries*, *spatial range queries*, and *spatial join queries* are frequently used in spatial database systems [26,19,7,23].

For rapid processing of spatial queries, the selection of an efficient indexing method is crucial. An excellent survey on spatial indexing methods can be found in [3]. They are classified into three categories as follows.

The first is a class of methods that divide the original space into a set of subspaces and maintain the relationship between those subspaces and their objects. R-trees [14,18], R*-trees [2], R$^+$-trees [33], and Cell-trees [12] are typical examples. We refer to this class of methods as *original space indexing methods* (OS-IMs) in this paper.

[*] Corresponding author. Tel./fax: +82 2 2220 4567.
E-mail addresses: backhyun@sysgate.co.kr (H. Back), jiwon@hanyang.ac.kr (J.-I. Won), jhyoon@hallym.ac.kr (J.-H. Yoon), sanghyun@cs.yonsei.ac.kr (S. Park), wook@hanyang.ac.kr (S.-W. Kim).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.11.029