PCR-Tree: An Enhanced Cache Conscious Multi-dimensional Index Structures

Young Soo Min¹, Chang Yong Yang¹, Jae Soo Yoo¹, Jeong Min Shim², and Seok Il Song³

¹ Department of Computer and Communication Engineering, Chungbuk National University
48 Gaesin-dong, Cheongju Chungbuk, Korea
{minys, cyyang, yjs}@netdb.chungbuk.ac.kr
² Electronics and Telecommunications Research Institute
161 Gajeong-dong, Yuseong-gu, Daejeon, Korea
jmshim@etri.re.kr
³ Department of Computer Engineering, Chungju National University
Iryu Meon Gumdan Lee, Chungju Chungbuk, Korea
sisong@chungju.ac.kr

Abstract. Recently, to relieve the performance degradation caused by the bottleneck between CPU and main memory, cache conscious multi-dimensional index structures have been proposed. The ultimate goal of them is to reduce the space for entries so as to widen index trees, and minimize the number of cache misses. They can be classified into two approaches according to their space reduction methods. One approach is to compress MBRs by quantizing coordinate values to the fixed number of bits. The other approach is to store only the sides of MBRs that are different from their parents. In this paper, we investigate the existing multi-dimensional index structures for main memory database systems through experiments under the various work loads. Then, we propose a new index structure that exploits the properties of the both techniques. We implement existing multi-dimensional index structures and the proposed index structure, and perform various experiments to show that our approach outperforms others.

1 Introduction

Recently, as the performance gap between CPU and main memory gets larger, it becomes increasingly important to consider cache behavior and to reduce L2 cache line misses for the performance improvement of MMDBMs [1], [2]. Subsequently, several researches to improve the performance of index structures for MMDBMSs by reducing L2 cache misses have been done actively in the database community [3], [4], [5], [6], [7], [8]. Since the end of 1990s, cache conscious index structures have been one of the primary concerns to improve the performance of MMDBMSs. Particularly, in the beginning of the 2000s, some cache conscious multi-dimensional index structures have been proposed to enhance the performance of modern applications such as
GIS (Geographical Information Systems) and LBS (Location Based Systems) based on MMDBMSs.

To our knowledge, most recent cache conscious multi-dimensional index structures are cache conscious R-tree (CR-tree)[7], partial R-tree (PR-tree)[8] and normal R-tree[9] with a small node size, a cache line size or its multiple. CR-trees compress MBRs by quantizing coordinate values to the fixed number of bits. PR-trees store only the sides of MBRs that are different from their parents. The ultimate goal of both index structures is to reduce the space for MBRs so as to widen index trees and minimize the number of cache misses.

In this paper, we investigate existing multi-dimensional index structures for MMDBMSs such as CR-trees, PR-trees and normal R-trees through some experiments. Then, we propose a new index structure that exploits the properties of CR-tree and PR-tree. Actually, the partial MBR method of PR-tree works well when the number of entries is small, generally 3~7. However, the number of entries of CR-tree that uses compression techniques is much more than that of normal R-tree. In that reason, partial MBR method seems to be inadequate for CR-tree. Through experiments, we investigate that integrating CR-tree with PR-tree is valuable. Finally, we perform extensive experiments to show that our proposed index structure outperforms CR-tree, PR-tree and R-tree in various environments.

This paper is organized as follows. In section 2, we describe existing cache conscious uni-dimensional and multi-dimensional index structure. In Section 3, we analyze PR-tree and CR-tree in detail, and propose PCR-tree in section 4. Section 5 presents experimental results that proposed algorithm outperforms existing index structures through various experiments. Finally, section 6 concludes this paper.

2 Related Works

In the end of 1990s, Rao and Ross proposed two index structures for MMDBMs that consider cache behavior. The index structures are Cache-Sensitive Search tree (CSS-tree)[3] and Cache-Sensitive B+-tree (CSB+-tree)[4]. CSS-tree designed for OLAP environments is a very compact and space-efficient B+-tree. It eliminates pointers of index trees completely, and stores keys in contiguous memory. CSB+-tree applies the idea of CSS-tree to an index structure for OLTP environments that should support efficient update. Partial key B+-tree (pkB+-tree) are proposed in [5]. The partial key approach of pkB+-tree uses fixed-size parts of keys and information about key differences to minimize the number of cache misses. [6] applies to prefetch techniques to B+-tree to eliminate the cache miss latency efficiently. The approach of pkB+/pB+-tree is orthogonal to that of CSS/CSB+-tree. Above methods are for uni-dimensional index structures such as B+-tree. Cache-conscious R-tree (CR-tree) and partial R-tree (PR-tree) are multi-dimensional index structures. We describe the features of the two index structures that form the basis of our proposed index structure.

In the PR-tree, usually the node size of R-trees in MMDBMSs is a cache line size (32 ~ 64 bytes) or the multiple of a cache line size. The authors of PR-trees performed an experiment to show how many MBRs in R-tree share their sides with their