

변환기법을 이용한 공간 액세스 구조의 영역분할 전략

윤동하, 이종학

대구가톨릭대학교 컴퓨터정보통신공학부
e-mail : jhlee11@cuth.cataegu.ac.kr

A Region Splitting Strategy for Spatial Access Structures Using Transformation Techniques

Dong-ha Yoon, Jong-Hak Lee

Dept. of Computer & Information Communications Engineering,
Catholic University of Taegu

요약

물리적 데이터베이스 설계기법은 최적의 질의처리 성능을 제공하기 위하여 데이터베이스의 액세스 구조를 결정하는 과정이다. 본 논문에서는 변환기법을 이용한 공간 액세스 구조의 물리적 데이터베이스의 설계를 위한 영역분할 전략을 제시한다. 변환기법을 이용한 공간 액세스 구조는 원공간(original space)에서의 공간 객체들을 공간의 차원을 두 배로 하는 변환공간(transformation space)내의 점 객체들로 변환하여 관리하는 방법이다. 먼저, 원공간에 주어지는 모든 공간 질의가 변환공간에서는 한가지 형태의 범위 질의로 변환되는 특징이 있음을 보인다. 그리고, 변환공간상에서 이 범의 질의가 위치하는 질의 영역의 모양과 데이터 페이지가 위치하는 페이지 영역의 모양 사이의 관련성을 이용하여 질의처리의 성능을 향상시킬 수 있는 영역분할 전략을 제안한다. 성능평가의 결과에 의하면, 주어진 질의 패턴에 따라 최적의 공간 액세스 구조를 구성할 수 있었으며, 이차원 원공간에 대한 사차원 변환공간인 경우에 질의의 형태에 따라 질의처리의 성능이 다섯배 이상까지 향상되었다.

1. 서론

지리 정보 시스템(GIS), VLSI 설계 등의 새로운 데이터베이스 응용에서는 공간상에서 크기를 가지는 객체인 공간 객체의 효율적인 관리가 필요하다. 최근들어 이를 위한 공간 데이터베이스 시스템에 대한 연구가 활발히 진행되고 있으며, 그 중에서도 공간 객체에 대한 데이터 모델링과 액세스 방법 분야에 연구가 집중되고 있다[6].

공간 데이터를 위한 액세스 방법으로 공간상에서 인접한 객체들을 한꺼번에 검색하는 등의 공간 질의 고유의 특성을 고려하는 새로운 액세스 방법인 공간 액세스 방법이 필요하다. 공간 액세스 방법은 공간 객체의 크기를 관리하는 방법에 따라 공간 채움 곡선 기법, 객체 분할 또는 중복 기법, 영역 접침 기법, 변환기법 등으로 분류할 수 있다[3].

변환기법[4]은 원공간(original space)에서의 공간 객체들을 공간의 차원이 두 배가 되는 변환공간(transform space)내의 점 객체들로 변환하여 관리함으로써 객체의 크기에 대한 별도의 관리를 없애기 위한 방법이다[5]. 대표적인 두 변환기법에는 구석점

변환과 중앙점 변환기법이 있다[4]. 이들 기법들은 포함자로서 객체를 최소한으로 포함하는 축에 수직인 선분들로 구성되는 최소 포함 사각형(Minimum Bounding Rectangle: MBR)을 이용하여 객체를 색인한다. 구석점 변환기법은 이차원 원공간에서 객체에 대한 MBR의 좌하점과 우상점의 네 좌표 값을 파라미터로 사용하고, 중앙점 변환기법은 객체에 대한 MBR의 중심점의 두 좌표값과 각 변 길이의 반을 네 파라미터로 사용하여 객체를 사차원 변환공간 내의 한 점으로 표현하는 것이다.

본 논문에서는 변환기법을 이용한 공간 액세스 구조의 물리적 데이터베이스의 설계를 위한 영역분할 전략을 제시한다. 물리적 데이터베이스 설계는 사전에 분석한 질의 정보를 기반으로 데이터베이스 파일의 클러스터링 특성과 액세스 방식 등을 결정함으로써 주어진 데이터베이스 응용의 질의처리 시 최적의 성능을 지원하는 방법이다[1]. 따라서, 공간 데이터베이스 시스템을 위한 액세스 구조에서도 이러한 물리적 데이터베이스 설계를 위한 액세스 구조의 영역분할 전략을 필요로 한다.

변환기법을 이용한 공간 액세스 구조에서는 변환 공간의 영역 분할전략을 통하여 변환공간을 구성하는 각 축별로 클러스터링의 정도를 조절할 수 있다. 기존의 공간 액세스 구조에서는 영역분할 전략으로 순환분할 전략(cyclic splitting strategy)[6]을 사용하여 모든 축들이 클러스터링 특성을 같은 정도로 공유하고 있다. 이것은 질의 조건에 나타나는 모든 축들의 구간 크기가 모두 같은 정방형의 질의 영역이 주로 사용된다는 것을 가정한 것으로, 각 축에 따라 질의 조건의 구간 크기가 다를 경우 질의처리 성능이 저하된다.

변환기법을 이용한 공간 액세스 구조에서는 원공간에 주어지는 모든 공간 질의가 변환공간에서는 한가지 형태의 범위 질의(range query)로 변환되는 특징이 있다. 변환공간상에서 이 범위 질의가 위치하는 영역을 질의 영역(query region)이라 할 때, 공간 데이터베이스 시스템의 사용자 질의는 질의 영역에 포함되는 객체들을 탐색하는 연산으로 해석할 수 있다. 따라서, 공간 액세스 구조의 데이터 페이지에 해당하는 변환공간상의 영역을 페이지 영역(page region)이라 할 때, 공간 액세스 구조의 물리적 데이터베이스 설계의 문제는 일련의 사용자 질의 영역들에 의해서 교차되는 페이지 영역들의 개수를 최소화하는 문제가 된다.

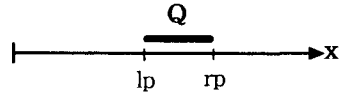
본 논문에서는 영역을 구성하는 각 축의 구간 크기에 대한 비율을 구간비(interval ratio)라 하고, 영역의 모양을 구간비로서 표현한다. 공간 사용자 질의 패턴에 나타나는 질의 영역들의 모양에 대한 정보를 기반으로 질의 영역들에 의해 교차되는 페이지 영역들의 개수가 최소로 되는 최적 페이지 영역의 구간비를 결정하고, 가능한 이와 같은 구간비를 갖는 페이지 영역들이 되도록 하는 영역 분할전략을 사용함으로써 최적의 공간 액세스 구조를 구성할 수 있다.

2. 공간 질의의 특징

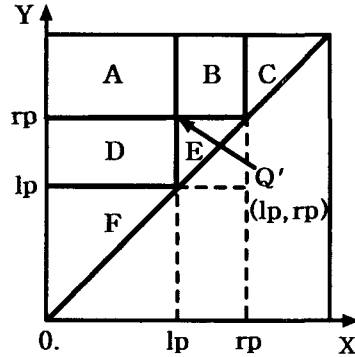
본 절에서는 변환기법을 이용한 공간 액세스 구조에서 원공간에 주어지는 모든 공간 질의가 변환공간에서는 한가지 형태의 범위 질의로 변환되는 특징이 있음을 알아본다. 공간 질의는 주어진 질의 영역에 대하여 특정 공간 관계를 가지는 객체들을 검색하는 질의이다. 이들 질의에는 영역 교차 질의, 영역 포함 질의, 영역 피포함 질의, 점 질의, 최근접 인접 질의 등의 기본 유형이 있다[5].

그림 1에서 (a)는 일차원 원공간의 한 질의 영역 Q를 나타내고, (b)는 이 질의 영역을 구석점 변환기법에 의한 이차원 변환공간상에 최측점 lp를 x축, 우측점 rp를 y축 값으로 나타낸 질의 점 Q' 과, 이를 중심으로 변환공간상에서 상대적 공간 관계에 따라 객체가 존재할 수 있는 영역을 A에서 F까지 여섯 영역으로 구분하여 나타낸 것이다. 예를들어, A는 원공간에서 Q를 포함하는 모든 객체들이 위치하는 영역이고, B는 원공간에서 Q의 우측점과 교차하는 모든 객체들이 위치하는 영역이며, C는 원공간에

서 Q의 우측에 존재하는 모든 객체들이 위치하는 영역이다.



(a) 원공간에서의 질의 영역 Q.



(b) 변환공간에서의 공간 관계별 영역 구분.

그림 1: 변환기법의 공간 관계별 영역 구분.

위와 같은 영역 구분을 이용하면 원공간의 다양한 영역 질의는 변환공간에서 한가지 형태의 범위 질의로 변환하여 처리할 수 있음을 알 수 있다. 즉, 다음은 주어진 질의 영역 Q에 대한 영역 질의들이 어떤 범위 질의로 변환되는가를 나타낸다.

- 영역 교차 질의: 그림 1(b)의 변환공간에서 검색되어야 하는 부분은 A, B, D와 E를 합한 영역이다. 따라서, 이 영역은 이차원 공간상에서 $x \leq rp$ AND $y \geq lp$ 인 범위 질의의 처리를 위하여 검색해야 하는 부분이다.
- 영역 포함 질의: 그림 1(b)의 변환공간에서 검색되어야 하는 부분은 영역 E이다. 따라서, 이 영역은 이차원 공간상에서 $lp \leq x \leq rp$ AND $lp \leq y \leq rp$ 인 범위 질의의 처리를 위하여 검색해야 하는 부분이다.
- 영역 피포함 질의: 그림 1(b)의 변환공간에서 검색되어야 하는 부분은 영역 A이다. 따라서, 이 영역은 이차원 공간상에서 $x \leq lp$ AND $y \geq rp$ 인 범위 질의의 처리를 위하여 검색해야 하는 부분이다.
- 점 질의: 주어진 질의 점 Qp에 대하여 이 질의를 위하여 검색할 부분은 이차원 변환공간상에서 $x \leq Qp$ AND $y \geq Qp$ 인 범위 질의의 처리를 위하여 검색해야 하는 부분이다.

3. 공간 액세스 구조의 영역분할 전략

참고문헌[2]에서는 이차원 파일구조를 구성하는 페이지 영역들의 최적 구간비를 다음과 같이 구할 수 있음을 밝혔다. 즉, 서로 다른 크기의 페이지 영역들로 나누어져 있는 이차원 공간상에서, 임의의

위치에 주어지는 n 개의 질의 영역 $q_i(x) \times q_i(y)$ ($i = 1, \dots, n$)에 대해 각 질의 영역의 객체 밀집도를 $d_i (= \frac{n_0}{q(x) \times q(y)})$, 단, n_0 는 질의 영역 내의 객체 수라 할 때, 각 질의 영역과 교차하게 되는 페이지 영역의 총 개수를 최소화 하는 페이지 영역의 최적 구간비($p(x) : p(y)$)는 $\sum_{i=1}^n q_i(x) \sqrt{d_i} : \sum_{i=1}^n q_i(y) \sqrt{d_i}$ 이다[2].

따라서, 본 논문에서는 먼저, 질의 패턴으로 주어지는 공간 질의들에 의해 다차원 변환공간상에 표현되는 질의 영역들로서 페이지 영역의 최적 구간비를 계산한다. 그리고, 공간 액세스 구조의 변환공간이 최적 구간비의 페이지 영역들로서 구성될 수 있도록 하는 영역분할 전략을 사용한다.

정리 1 구간비가 $q_x : q_y$ 인 $q_x \times q_y$ 형태의 질의 영역이 이차원 공간상에서 임의의 위치에 주어질 때, 크기가 B 인 $p(x) \times p(y)$ 형태의 한 페이지 영역과 교차하게 되는 질의 영역의 위치 범위의 크기는 페이지 영역의 구간비가 주어진 질의 영역의 구간비와 같을 때 최소가 된다.

증명: 아래 그림 2는 이차원 도메인 공간에서 $q_x \times q_y$ 형태의 질의 영역 Q 가 임의의 위치에 주어질 때, 크기가 $B (= p(x) \times p(y))$ 인 특정 페이지 영역 P 와 교차하게 되는 위치의 범위를 질의 영역 Q 의 좌상점이 위치할 수 있는 영역(빗금친 부분) LQ 로 나타낸 것이다.

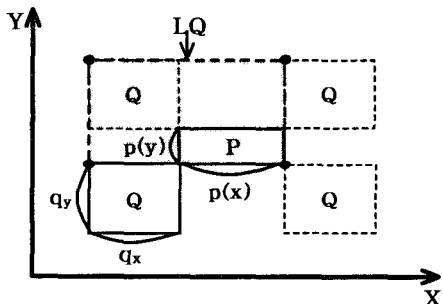


그림 2: 임의의 한 페이지 영역과 교차하게 되는 질의 영역의 위치 범위.

그림 2에서 LQ 의 크기 $SIZE_LQ(p(x), p(y))$ 는 다음 식과 같다.

$$SIZE_LQ(p(x), p(y)) = (p(x) + q_x)(p(y) + q_y) \quad (1)$$

$p(x) \times p(y) = B$ 에 의해서, 수식(1)에서 $p(y)$ 를 $\frac{B}{p(x)}$ 로 치환하여 수식(1)의 값을 최소화 하는 $p(x)$ 를 구하면, $p(x) = \sqrt{(q_x/q_y)B}$ 이다. 또한, 이러한 $p(x)$ 에 대한 $p(y)$ 는 $p(x) \times p(y) = B$ 에 의하여 $p(y) = \sqrt{(q_y/q_x)B}$ 이다. 그러므로, $SIZE_LQ(p(x),$

$p(y))$ 를 최소로 하는 페이지 영역 P 의 구간비 $p(x):p(y) = q_x:q_y$ 이다. □

정리 1을 이용하여 페이지 영역의 분할 시 분할된 페이지 영역의 구간비가 최적 구간비에 가깝게 되는 분할 축을 선택할 수 있다. 그림 3은 주어진 최적 구간비 ($a:b$)와 같은 모양을 갖는 $a \times b$ 형태의 질의 영역 Q 가 이차원 도메인 공간상에 임의의 위치에 주어졌다고 가정하고, $p(x) \times p(y)$ 형태의 페이지 영역 P 가 두 개의 페이지 영역으로 분할된 후의 한 페이지 영역과 교차하게 되는 질의 영역의 위치 범위(빗금친 부분) LQ 를 나타낸다.

그림 3에서 X축을 분할했을 경우의 LQ_x 의 크기 $SIZE(LQ_x)$ 는 $(p(x)/2 + a)(p(y) + b)$ 이고, Y축을 분할했을 경우의 LQ_y 의 크기 $SIZE(LQ_y)$ 는 $(p(x) + a)(p(y)/2 + b)$ 이다.

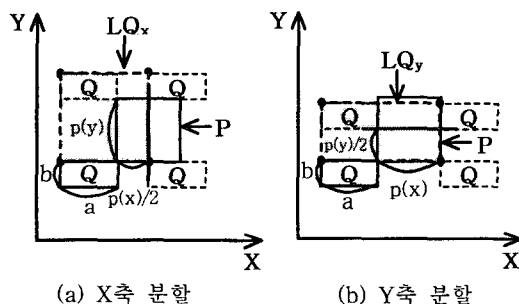


그림 3: 분할후의 한 페이지 영역과 교차하게 되는 질의 영역의 위치 범위.

정리 1에 의하여 LQ 의 크기는 페이지 영역의 구간비가 질의 영역의 구간비와 같을 때 최소가 되므로, 이 LQ 의 크기가 작을수록 페이지 영역의 구간비가 질의 영역 Q 의 구간비에 근접하게 된다. 따라서, 이 LQ 의 크기가 작게 되는 축을 분할 축으로 선택함으로써 분할 후의 페이지 영역의 구간비를 주어진 최적 구간비에 접근하게 할 수 있다.

이와같은 이차원 변환공간의 영역분할 전략을 사차원 변환공간으로 확장하여 적용할 수 있다. 즉, W, X, Y, Z 축으로 이루어진 사차원 변환공간에서 페이지 영역의 분할 시 분할된 페이지 영역의 구간비가 주어진 최적 구간비에 가장 가깝게 되는 분할 축을 선택할 수 있다. 먼저, 최적 구간비 ($a:b:c:d$)와 같은 모양의 질의 영역 $Q(a \times b \times c \times d)$ 가 사차원 변환공간상에 임의의 위치에 주어진다고 가정하고, 분할이 요구되는 페이지 영역 ($p(w) \times p(x) \times p(y) \times p(z)$)이 각 축에 대해 분할된 후의 한 페이지 영역과 교차하게 되는 질의 영역의 위치 범위 LQ 의 크기를 계산하여, 그 값이 가장 작게 되는 축을 분할 축으로 선택하면 된다. 예를들어, 분할 축으로 W 축을 선택했을 때 LQ_w 의 크기 $SIZE(LQ_w)$ 는 $(p(w)/2 + a)(p(x) + b)(p(y) + c)(p(z) + d)$ 이다.

4. 성능 평가

본 절에서는 제안한 영역분할 전략의 유용성을 검증하기 위하여 변환기법을 이용한 공간 액세스 구조의 하나인 MBR-계층 그리드 파일(MBR-Multilevel Grid File: MBR-MLGF)[7]로서 실험한 결과를 제시한다.

본 실험에서는 사차원의 비균일 분포를 가지는 데이터에 대하여 특정 구간비의 페이지 영역을 갖는 여러 개의 사차원 MBR-MLGF들을 생성하고, 각각에 대하여 고유의 질의 영역 형태를 갖는 여러 질의들로 구성된 질의패턴을 처리할 때 발생하는 페이지 액세스 수를 측정하였다. MBR-MLGF들의 구축에 사용된 페이지 영역의 구간비는 각각 1:1:1:1, 1:2:4:8, 1:4:16:64, 1:8:64:512, 및 1:16:256:4096 등 다섯 가지이며, 질의 영역의 경우에는 크기가 도메인 공간의 1/20000로서 소영역인 경우에만 한정하여 질의 영역의 구간비가 각각 1:1:1:1, 1:2:4:8, 1:4:16:64, 1:8:64:512, 및 1:16:256:4096인 S1_1_1_1, S1_2_4_8, S1_4_16_64, 및 S1_8_64_512, 및 S1_16_256_4096 형태의 질의 영역 등이다.

각 질의 영역의 형태별로 1000개의 질의 영역을 도메인 공간상에 균일하게 생성하고, 이들 질의를 처리하는데 발생하는 평균 페이지 액세스 수를 측정한다. 그림 4는 이에 대한 실험 결과를 그래프로 나타낸 것이다. 모든 형태의 질의 영역에 대하여, 그 질의 영역의 구간비를 페이지 영역의 구간비로 가지는 MBR-MLGF에서 가장 좋은 성능을 보인다. 그리고, 그림 4에서 나타난 바와 같이 질의 영역의 구간비가 1:16:256:4096인 경우 페이지 영역의 구간비가 기존의 1:1:1:1인 MBR-MLGF에 비해 페이지 영역의 구간비가 1:16:256:4096인 MBR-MLGF에서 질의처리 성능이 다섯배 이상까지 향상됨을 볼 수 있다. 이와같은 실험 결과는 본 논문에서 제안한 공간 액세스 구조의 영역분할 전략이 매우 유효함을 보이는 것이다.

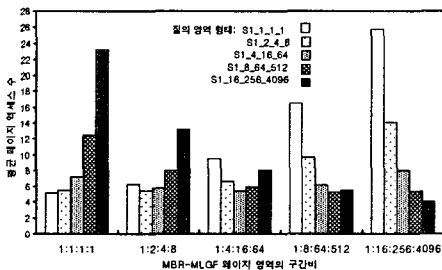


그림 4: 다양한 MBR-MLGF별 질의 영역의 형태별 질의 처리 성능.

5. 결론

본 논문에서는 변환기법을 이용한 공간 액세스 구조에 대하여, 먼저 원공간의 다양한 공간 질의들은 변환공간에서는 한 가지 형태의 범위 질의들로

변환하여 처리할 수 있음을 보이고, 이러한 범위 질의들을 처리하기 위하여 액세스하는 데이터 페이지의 개수를 최소로 하는 영역분할 전략을 제시한다. 제안한 영역분할 전략은 먼저, 도메인 공간상의 데이터 분포 특성을 고려한 질의 패턴에 대한 정보로서 데이터 페이지 영역의 최적 구간비를 구하고, 공간 액세스 구조의 페이지 영역의 모양이 이 최적 구간비에 최대한 근접하도록 하는 영역분할 전략이다.

제안한 영역분할 전략의 성능평가를 위한 실험 결과에 의하면, 주어진 질의 패턴과 데이터 분포에 따라 최적의 공간 액세스 구조를 구성할 수 있으며, 이차원 원공간에 대한 사차원 변환공간상의 질의 영역의 모양이 편향된 정도에 따라 기존의 정방형(구간비가 1:1:1:1인 경우) 모양의 페이지 영역으로 구성된 공간 액세스 구조에 비해 질의처리의 성능이 매우 크게 향상되었다. 특히, 질의 영역의 구간비가 1:16:256:4096인 경우에는 질의처리 성능이 다섯배 이상까지 향상됨을 볼 수 있었다.

참고문헌

- [1] Elmasri, R. and Navathe, S. B., *Fundamentals of Database Systems*, Benjamin/Cummings Publishing Co., Redwood City, California, Third Ed., 2000.
- [2] Lee, J. H. et al., "A Physical Database Design Method for Multidimensional File Organizations," *Information Sciences*, Vol. 102, No. 3, pp. 31-65, 1997.
- [3] Lu, H. and Ooi, B. C., "Spatial Indexing: Past and Future," *IEEE Data Engineering Bulletin*, Vol. 16, No. 3, pp. 16-21, Sept. 1993.
- [4] Pagel, B. U. et al., "The Transformation Technique for Spatial Objects Revisited," In *Proc. 3rd int'l Symp. on Spatial Databases(SSD'93)*, 1993.
- [5] Seeger, B. and Kriegel, H. P., "Techniques for Design and Implementation of Efficient Spatial Access Methods," In *Proc. 14th int'l Conf. on Very Large Data Bases*, pp. 360-371, 1988.
- [6] Song, J. W., Whang, K. Y., and Kim, S. W., "Spatial Join Processing Using Corner Transformation," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 11, No. 4, Aug. 1999.
- [7] Whang, K. Y., Kim, S. W., and Wiederhold, G., "Dynamic Maintenance of Data Distribution for Selectivity Estimation," *The VLDB Journal*, Vol. 3, No. 1, pp. 29-51, Jan. 1994.