

고차원 공간 데이터를 위한 연속 범위 질의의 효율적인 처리

장수민[○] 유재수

충북대학교 정보통신공학과

jsm[○]@netdb.chungbuk.ac.kr, yjs@chungbuk.ac.kr

An Efficient Processing of Continuous Range Queries on High-Dimensional Spatial Data

Sumin Jang[○] Jaesoo Yoo

Dept of Computer and Communication Engineering, ChungBuk National University

이동 객체에 대한 연속 범위 질의(Continuous Range Query)의 응용프로그램이 급속도로 확장되면서 이차원 정보를 넘어서 고차원 공간 데이터에 대한 처리를 요구하고 있다. 기존의 색인 기법들은 대용량의 고차원 공간 데이터에 대한 연속 범위 질의를 처리 시 객체의 수와 질의의 수가 증가함에 따라 질의 처리하기 위한 저장 비용과 처리 성능이 저하되는 문제점이 있다. 또한 질의 속성이 2차원이 아닌 고차원임으로 성능저하가 심하다. 본 논문은 이러한 문제점을 해결하기 위하여 고차원 공간데이터에 대하여 저장 비용이 적고 효율적인 질의 처리 시간을 제공하는 PAB (Projected Attribute Bit)-기반의 질의색인 방법을 제안한다. 또한 제안하는 질의 색인 기법은 효율적인 질의 처리를 위하여 점진적인 갱신(Incremental Update)을 지원한다.

공간데이터에 대한 색인 기법으로는 크게 R-트리계열 방식과 그리드계열 방식으로 분리된다. R-트리계열 방식은 디스크기반이 아닌 메모리기반으로 변경하여도 그리드계열 방식보다 성능이 좋지 않다. 그리드 계열 방식으로 VCR(Virtual Construct Rectangles)-기반의 질의색인[2]와 CES (Containment Encoded Squares)-기반의 질의색인[1]은 최근의 질의색인기법들이다. 그러나 이러한 그리드 계열 방식은 부분적으로 중첩되는 질의들이 많아지고 특히 차원이 높아질수록 저장비용이나 처리시간이 지속적으로 증가하기 때문에 고차원 공간데이터를 처리하기에 적합하지 않다.

제안하는 방법은 이러한 투영이라는 작업을 통하여 연속 범위 질의의 고차원 정보를 일차원 정보들로 변환한다. 삼차원 공간의 입체도형을 특정한 규약을 가지고 이차원 평면상에 나타내는 방법으로 삼차원 데이터를 이차원으로 차원을 한 차원 낮추어 표현하는 방법을 투영이라 한다. 그림 1은 3차원 공간의 질의를 1차원 정보로 변환하기 위한 투영과정이다. 그림 1(a)과 같이 질의는 직육면체로 표현되고 이동객체는 3차원 공간에 위치한다. 이러한 직육면체 질의를 그림 1(b)과 같이 질의 q 에 대하여 빛 $La1$ 을 수직으로 비추면 $a1$ 축에 그림자 $SDa1(q)$ 을 생성한다. 그림자 $SDa1(q)$ 은 $a1$ 축에 생성된 질의 q 의 투영도가 된다. 각 축마다 이러한 투영작업을 하여 투영도를 생성하고 투영도들을 그리드형태로 분할하여 그림자가 있는 부분은 1로 그림자가 없는 부분은 0으로 설정한다. 이러한 비트 리스트를 PQBL(Projected Query Bit List)라 명칭하고 분할된 한 개의 셀은 PAB이다. 이러한 질의의 속성을 내포한 PQBL을 통하여 어떤 객체가 어떤 질의에 포함되는지를 검색한다. 만약 객체가 3개의 그림자 $SDa1(q)$, $SDa2(q)$, $SDa3(q)$ 의 범위 안에 있다면 그 객체는 범위질의 q 에 포함된다.

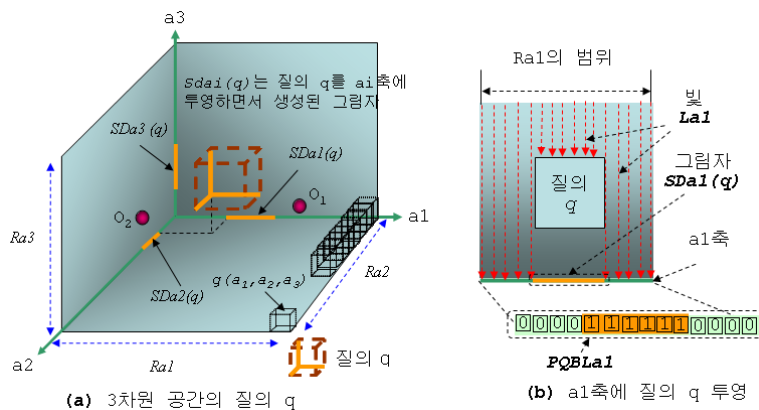


그림 1. 3차원 공간에서 질의를 투영하는 과정

*본 연구는 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단과 산업자원부 지역혁신인력양성사업의 연구결과물입니다.

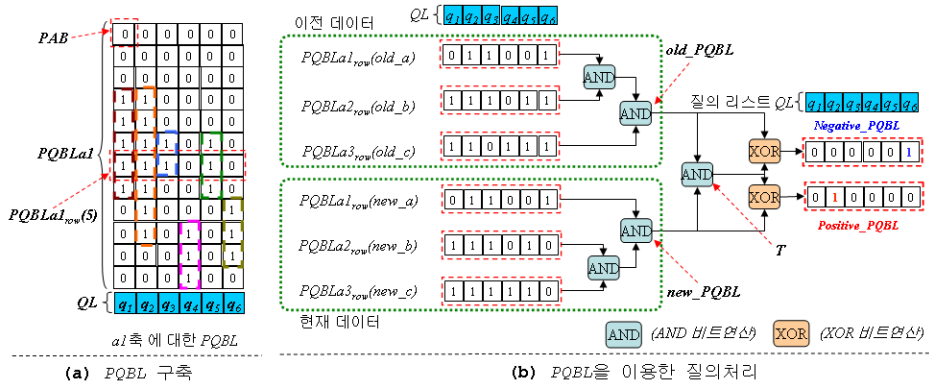
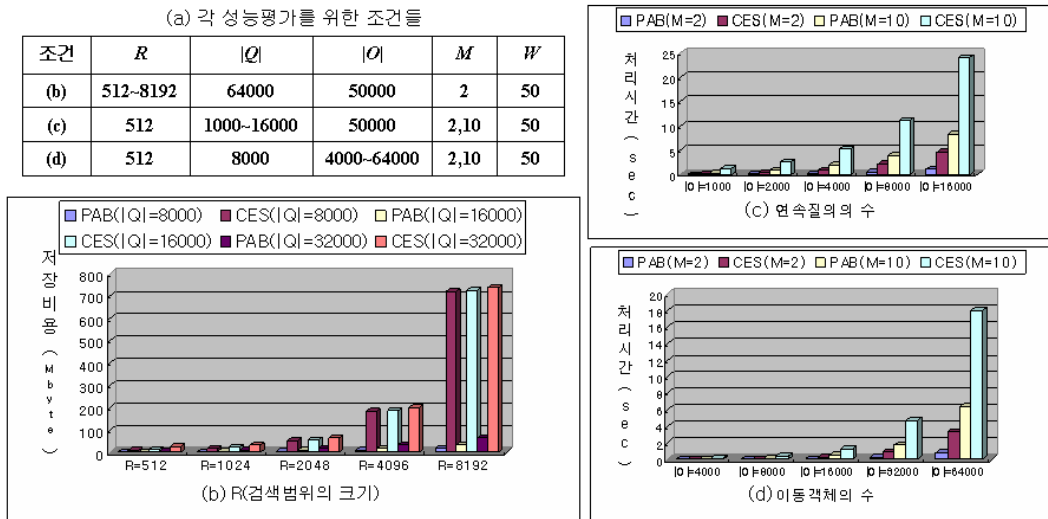


그림 2. 3차원 데이터에 대한 질의처리 과정

질의처리과정은 다음과 같다. 첫 번째로, 그 객체의 이전 위치에 해당하는 3개 $PQBL_{row}$ 들과 그 객체의 새로운 위치에 해당하는 3개 $PQBL_{row}$ 들을 얻는다. 두 번째로, 이전 위치와 새로운 위치에 해당하는 각각의 3개의 $PQBL_{row}$ 들을 AND 비트 연산을 하여 old_PQBL 와 new_PQBL 를 구한다. 이러한 old_PQBL 와 new_PQBL 는 그 객체가 이전 위치나 새로운 위치에 있을 때 포함되는 질의들의 리스트에 대한 정보를 포함한다. 또한 old_PQBL 와 new_PQBL 의 AND 비트 연산을 통하여 T 을 만든다. 세 번째로, old_PQBL 와 T 의 XOR 비트 연산을 통하여 $Negative_PQBL$ 을 만들고 new_PQBL 와 T 의 XOR 비트 연산을 통하여 $Positive_PQBL$ 을 만든다. 이러한 과정을 통하여 만들어진 $Negative_PQBL$ 는 질의의 이전결과에 그 객체를 삭제하기 위한 질의의 리스트이다 반대로, $Positive_PQBL$ 는 질의의 이전결과에 그 객체를 새로 삽입하기 위한 질의의 리스트이다. 이처럼 제안하는 기법은 빠른 질의처리를 위해 점진적으로 갱신하는 방식을 사용한다. 기존 질의색인기법들은 고차원의 데이터를 처리하기 위하여 많은 처리시간과 저장 비용을 필요로 한다. 그러나 제안하는 방식은 단지 추가 되어진 차원의 수만큼 $PQBL$ 을 추가하고 $PQBL$ 들간의 비트 연산을 통하여 질의 처리한다.

성능평가는 최근에 발표된 CES-기반의 질의색인기법과 비교분석하였다. 평가과정에서는 질의색인은 변화하지 않는다고 가정하고 CES-기반의 질의색인의 L 의 값은 실험에서 가장 좋은 결과를 산출하는 값으로 설정하였다. 연속 범위 질의는 2차원 공간에 직사각형으로 표현하고 질의의 크기는 W 로 무작위로 각각 선택되었다. 이동객체의 새로운 위치는 이전의 위치를 기준으로 임의의 방향으로 움직이며 최대이동크기 M 을 사용한다. 성능평가는 주기억 장치 1024 Mbytes로 Windows XP의 3.0GHz 펜티엄 IV 시스템에서 실행되었다. 그림 3(a)는 각각의 성능평가를 위한 조건들의 값을 설정한 것이다. 그림 3(b)는 검색영역크기 R 과 질의의 수 $|Q|$ 의 변화에 따른 색인저장 비용을 비교하였다. 그림 3(c)는 질의의 수 $|Q|$ 와 이동 객체의 이동크기 M 의 변화에 따른 질의처리 시간을 산출하였고 그림 3(d)는 객체의 수 $|O|$ 와 이동 객체의 이동크기 M 의 변화에 따른 질의처리 시간 성능평가 하였다. 제안하는 PAB -기반의 질의색인 기법은 고차원 공간데이터에 대한 질의색인비용과 질의처리시간 모든 측면에서 효율적임을 입증하였다.



참고문헌

[1] K.-L. Wu, S.-K. Chen, and P.S. Yu, "Incremental Processing of Continual Range Queries over Moving Objects". In IEEE TKDE, Vol.18, pp.1560-1575, 2006.
 [2] K.-L.Wu, S.-K. Chen, and P.S. Yu, "Processing continual range queries over moving objects using VCR-based query indexes". In Proc on Mobile and Ubiquitous Systems, 2004