

고차원 공간 데이터를 위한 연속 범위 질의의 효율적인 처리

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

장수민^{*} 유재수^{**}
(Sumin Jang) (Jaesoo Yoo)

요약 이동객체에 대한 연속 범위 질의(Continuous Range Query)의 응용프로그램이 급속도로 확장되면서 이차원정보를 넘어서 고차원 공간 데이터에 대한 처리를 요구하고 있다. 만약 고차원 데이터에 대한 중첩되어지는 연속 범위 질의의 정보를 기존의 색인으로 구성한다면 객체의 수와 질의의 수가 증가함에 따라 질의처리성능이 저하된다. 본 논문은 이러한 문제점을 해결하기 위하여 PAB(Projected Attribute Bit)-기반의 질의색인방법을 제안한다. 제안하는 기법은 성능향상을 위하여 질의의 정보를 각 속성 축에 투영이라는 작업을 통하여 고차원의 데이터를 1차원 정보들로 변환하고 이러한 정보를 비트단위로 구성하였다. 또한 제안하는 질의색인은 보다 효율적인 질의 처리를 위하여 점진적인 갱신(Incremental Update)을 지원한다. 다양한 성능평가 및 분석을 통하여 제안하는 방법이 최근에 연구된 CES-기반의 질의색인 기법보다 더 나은 확장성(Scalability)을 가짐을 입증한다.

키워드 : 질의색인, 위치기반서비스, 연속범위질의

* 본 연구는 2006년도 정부(과학기술부)의 재원으로 한국과학재단의 지원(No. R01-2006-000-1080900)과 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

· 이 논문은 2007 한국컴퓨터종합학술대회에서 '고차원 공간 데이터를 위한 연속 범위 질의의 효율적인 처리'의 제목으로 발표된 논문을 확장한 것임

^{*} 학생회원 : 충북대학교 정보통신공학과
jsm@netdb.chungbuk.ac.kr

^{**} 종신회원 : 충북대학교 정보통신공학과 교수
yjs@chungbuk.ac.kr
(Corresponding author)

논문접수 : 2007년 9월 27일

심사완료 : 2007년 11월 8일

: 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨터의 실제 및 레터 제13권 제6호(2007.11)

Copyright©2007 한국정보과학회

Abstract Recent applications on continuous queries on moving objects are extended quickly to various parts. These applications need not only 2-dimensional space data but also high-dimensional space data. If we use previous index for overlapped continuous range queries on high-dimensional space data, as the number of continuous range queries on a large number of moving objects becomes larger, their performance degrades significantly. We focus on stationary queries, non-exponential increase of storage cost and efficient processing time for large data sets. In this paper, to solve these problems, we present a novel query indexing method, denoted as PAB (Projected Attribute Bit)-based query index. We transfer information of high-dimensional continuous range query on each axis into one-dimensional bit lists by projecting technique. Also proposed query index supports incremental update for efficient query processing. Through various experiments, we show that our method outperforms the CES(containment-encoded squares)-based indexing method which is one of the most recent research.

Key words : query indexing, location-aware application, continuous range query

1. 서론

무선 통신망 시스템의 발전에 따라 단순히 데이터 통신 서비스만이 아닌 위치기반서비스(Location-Aware Service)의 비중이 증가하고 있다. 이처럼 비중이 확대 되는 위치기반서비스에서 연속 범위 질의(Continuous Range Query)의 처리는 매우 중요한 질의처리로 많은 연구가 되었다. 이동객체의 위치를 기반으로 하는 색인은 객체가 이동할 때 마다 모든 질의에 대하여 재평가와 색인을 수정해야 하는 단점이 있다. 그래서 질의의 정보를 기반으로 하는 질의색인을 구성하여 처리한다. [1]에서는 질의색인이 객체색인보다 우수함을 입증하였다. 그러나 기존의 질의색인방법들도 고차원 공간데이터를 처리하는데 높은 저장비용과 많은 처리시간을 필요로 하는 문제점이 있다.

이동객체에 대한 연속 범위 질의의 처리방법은 대략적으로 크게 두 가지의 종류로 분류한다. 첫 번째는 이동객체에 대하여 위치가 정적인 연속질의(Continuous Stationary Query)를 처리하는 방법이다[1-5]. 두 번째는 이동객체에 대하여 위치가 동적인 연속질의(Continuous Moving Query)를 처리하는 방법이다[6-8]. 본 논문은 정적인 연속질의를 처리하는 방법에 해당된다.

Grid 질의색인[4], VCR기반의 질의 색인[3], CES-기반의 질의색인[9]는 정적인 연속 범위 질의를 처리하는 기법들이다. 이러한 기법들은 그리드 계열의 질의색인이

다 그러나 질의와 객체의 수가 증가하면 저장비용이 급격히 증가한다.

본 논문에서는 이러한 문제점을 해결하기 위하여 새로운 PAB(Projected Attribute Bit)-기반의 질의색인 기법을 제안한다. PAB-기반의 질의색인방법은 투영이라는 방법을 통하여 n차원의 질의정보를 각 속성(Attribute)별로 분리하여 색인을 구성한다. 제안하는 질의색인은 보다 효율적인 질의처리를 위해서 Negative 비트 리스트와 Positive 비트 리스트를 통하여 점진적 업데이트(Incremental Update)를 제공한다.

본 논문의 구성은 2장에서 관련연구에 대하여 기술하고, 3장에서는 PAB-기반의 질의색인을 제안한다. 4장에서는 제안하는 방법의 우수성을 입증하기 위해 성능평가와 분석결과를 기술한다. 마지막으로 5장에서는 본 논문의 결론과 향후연구방향을 제시한다.

2. 관련 연구

전통적인 공간색인방법은 R-Tree[10]과 같이 디스크 기반 방식을 사용한다. 그러나 디스크기반의 방식은 질의처리시간이 늦기 때문에 효과적이지 않다. [5]에서는 그리드를 이용한 색인이 R-tree 계열의 색인보다 성능이 향상되었으며 이동객체의 위치정보를 기반으로 하는 객체색인보다 질의의 정보를 기반으로 하는 질의색인이 보다 좋은 성능을 보여주었다.

VCR(Virtual Construct Rectangles)-기반의 질의색인[3]과 CES(Containment-Encoded Squares)-기반의 질의색인[9]는 최근의 연구된 그리드 계열의 질의색인기

법들이다. VCR-기반의 질의색인은 VCR를 포함하는 범위질의와 활성화된 VCR 사이에 직접적인 매핑을 통하여 유지한다. VCR은 지붕의 기와장이나 벽면을 채우는 타일과 같은 것이다. VCR들로 구성된 하나의 집합은 미리 정해진 유일한 ID로 가지도록 미리 정의된다. 그리고 질의의 영역은 하나나 그이상의 VCR들로 채워진다. 이 방법은 질의의 영역을 맡은 VCR들을 사용하여 임의의 위치에 있는 이동객체에 대한 결과를 선 처리함으로써 비용적인 면에서 효과적이다.

CES-기반의 질의색인은 VCR-기반의 질의색인과 마찬가지로 CES는 유일한 ID로 미리 정의한다. CES들은 각 질의의 범위를 맡기 위한 VC(Virtual Constructs)들이고 간접적으로 미리 처리된 질의의 결과를 저장하기 위하여 사용된다. CES-기반의 질의색인은 성능평가를 통하여 VCR-기반의 질의색인보다 효율적인 결과를 보여준다.

3. PAB-기반의 질의색인

본 논문에서는 위치기반서비스를 위한 빠른 처리시간과 낮은 저장비용을 지원하는 PAB-기반의 질의색인을 제안한다. 표 1은 이 논문에서 사용된 중요한 표기법 및 정의를 요약한 것이다.

3.1 질의의 투영

그림 1은 연속 범위 질의의 이차원 정보를 일차원 정보로 변형시키는 과정을 보여준다. 질의 q_i 에 대하여 빛 $L(A_x)$ 을 수직으로 비추면, A_x 의 축에 그림자 $SD(A_x, q_i)$ 를 생성하게 된다. 또한 같은 방법으로 그림자 $SD(A_y,$

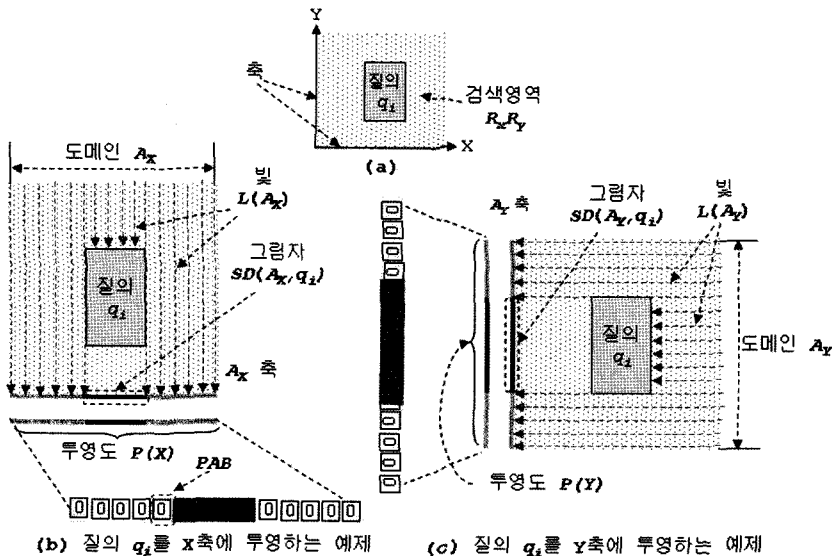


그림 1 연속 범위 질의를 투영하는 과정

표 1 기호에 대한 정의

표시	정의
$g(x,y)$	검색범위의 최소단위 $((x,y)$ 은 grid 좌표이다.)
$PQBL$ $PQBL(A)$	투영된 질의의 비트 리스트 (A 는 질의의 속성이다.)
$PQBL_{col}$ $PQBL_{col}(A, q_i)$	투영된 질의의 비트 리스트의 행 (A 는 질의의 속성이고, q_i 은 질의를 위한 ID 이다.)
$PQBL_{row}$ $PQBL_{row}(A,i)$	투영된 질의의 비트 리스트의 열 (A 는 질의의 속성이고, i 는 grid 좌표이다.)
$QL, QL(o_i)$	질의 ID 리스트(o_i 는 객체)

q_i)를 생성한다. 이때 이러한 투영작업을 통해서 각각의 그림자정보를 내포하는 두 개의 투영도($P(X), P(Y)$)를 얻는다. 그 그림자 정보를 저장하기 위해 검색범위의 최소단위 g 단위로 분할하고 그림자가 있는 곳은 1의 값으로 없는 부분은 0 값으로 설정한다. 분할된 한 개의 셀이 본 논문에서 언급하는 PAB이다. PAB은 1이나 0의 값을 갖는 비트로 구성된다.

3.2 PQBL 구성

PAB-기반의 질의색인은 연속 범위 질의의 속성 수만큼 PAB들로 이루어져 있는 PQBL들로 구성된다. X 속성에 대한 $PQBL(X)$ 는 범위질의 수만큼 $PQBL_{col}$ 들로 구성되고 R_x 의 도메인크기만큼 $PQBL_{row}$ 들로 구성된다. i 번째의 $PQBL_{col}$ 는 QL 의 i 번째 질의에 대응되고 X 속성에 대한 $PQBL_{col}(X, q_i)$ 는 QL 의 q_i 에 대응한다. $PQBL(X)$ 의 PAB들은 다음과 같은 절차에 의해 설정

된다. 첫째로, $PQBL_{col}(X, q_i)$ 의 PAB 리스트는 모두 0의 값으로 초기화한다. 둘째로, $PQBL_{col}(X, q_i)$ 의 비트 리스트에 범위질의 q_i 을 투영하여 그림자를 만든다. 마지막으로, 투영작업을 통하여 생긴 그림자의 위치에 대응되는 $PQBL_{col}(X, q_i)$ 의 PAB들은 1의 값으로 설정한다. $PQBL(Y)$ 는 $PQBL(X)$ 와 동일한 절차에 의해 구성된다.

그림 2는 PAB-기반의 질의색인을 위한 PQBL들을 구성하는 과정을 예제를 통하여 보여준다. 질의 q_i 을 위한 PAB들의 값은 다음과 같이 지정된다. 첫째로, q_i 의 $PQBL_{col}(X, q_i)$ 는 0의 값으로 초기화 한다. 둘째로, $PQBL_{col}(X, q_i)$ 에 질의 q_i 를 투영한다. 셋째로, q_i 의 투영으로 생긴 그림자가 있는 $PQBL_{col}(X, q_i)$ 의 두 개의 PAB(여섯, 일곱 번째)은 1의 값으로 채우면 된다.

3.3 질의 검색

그림 3은 PAB-기반의 질의색인을 사용하여 질의를 처리하는 알고리즘에 대한 의사코드를 보여준다. 처리과정은 첫째로, 그 객체의 이전 위치에 해당하는 2개 $PQBL_{row}$ 들과 그 객체의 새로운 위치에 해당하는 2개 $PQBL_{row}$ 들을 얻는다.

둘째로, 이전 위치와 새로운 위치에 해당하는 각각의 2개의 $PQBL_{row}$ 들을 AND 비트연산을 하여 $And_oldPQBL$ 와 $And_newPQBL$ 를 구한다. 이러한 $And_oldPQBL$ 또는 $And_newPQBL$ 는 그 객체가 이전 위치나 새로운 위치에 있을 때 포함되는 질의들의 리스트에 대한 정보를 포함하고 있다.

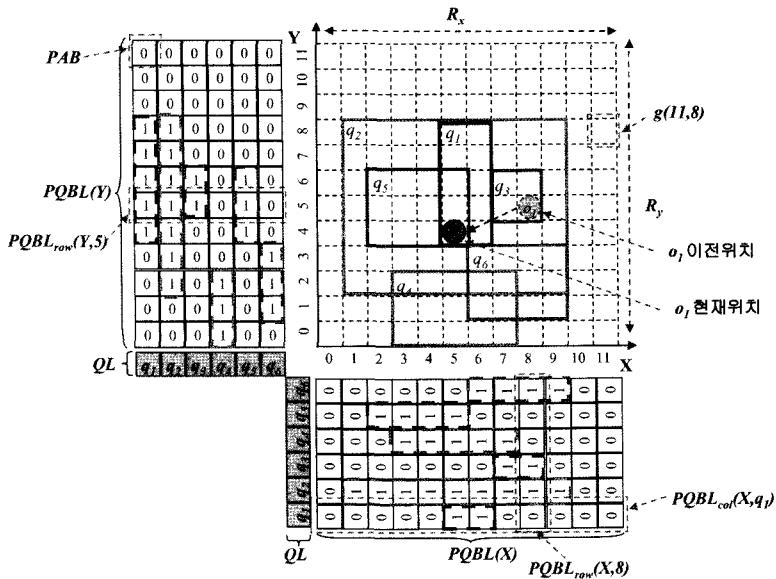


그림 2 PQBL를 만드는 과정

```

For (j=0;  $a_j \in O$  j++) {
  If(  $a_j$  (oldx,oldy) ==  $a_j$  (newx,newy) ) {
    continue;
  } else {
    // 이전 위치를 위한 AND 비트연산
    And_oldPQBL=PQBL_row(X, oldx) & PQBL_row(Y, oldy)
    // 현 위치를 위한 AND 비트연산
    And_newPQBL=PQBL_row(X, newx) & PQBL_row(Y, newy)
    And_PQBL= And_oldPQBL & And_newPQBL
    //negative 질의 업데이트를 위한 XOR 비트연산
    Negative_PQBL= And_PQBL ^ And_oldPQBL
    //positive 질의 업데이트를 위한 XOR 비트연산
    Positive_PQBL= And_PQBL ^ And_newPQBL
    If(Negative_PQBL > 0) {
      // nq 는 Negative_PQBL의 비트 리스트의 값이 1에
      // 해당하는 질의이다.
      Delete  $a_j$  from nq
    }
    If(Positive_PQBL > 0) {
      // pq 는 Positive_PQBL의 비트 리스트의 값이 1에
      // 해당하는 질의이다.
      Insert  $a_j$  to pq ;
    }
  }
}
    
```

그림 3 질의처리의 의사코드

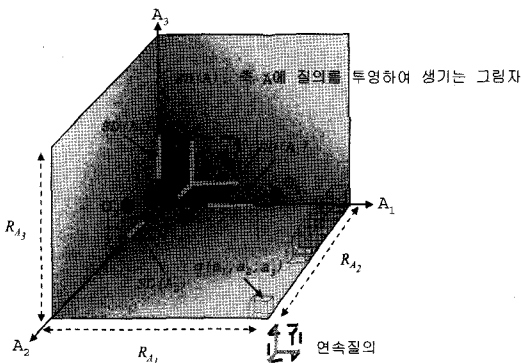


그림 4 3차원 데이터를 위한 투영작업

3.4 고차원 데이터를 위한 질의처리

제한하는 방식의 경우에는, 차원의 수만큼 PQBL을 추가하면 된다. 그림 4는 삼차원공간에서 질의처리를 위한 세 개의 PQBL들을 만들기 위하여, 연속 범위 질의는 각 속성의 축 (A_1, A_2, A_3)에 연속 범위 질의를 투영하여 3개의 그림자($SD(A_1), SD(A_2), SD(A_3)$)를 생성하고 그 그림자들의 정보에 의해 3개 PQBL들은 구성하는 과정을 보여준다.

4. 성능평가

성능평가는 CES-기반의 질의색인과 PAB-기반의 질의색인을 서로 비교분석하였다. 본 성능평가에서는 질의 색인이 변화하지 않는다고 가정하고 CES-기반의 질의 색인의 L의 값은 실험에서 가장 좋은 결과를 나타내는 값으로 설정된다. 질의의 크기는 1과 W 사이의 값으로 무작위로 각각 선택하였다. 이동객체는 이전의 위치를

기준으로 임의의 방향으로 이동하도록 설정하였다. 객체의 이동단위는 최대이동크기 M을 사용한다. 성능평가는 주기의 장치 1024 Mbytes로 Windows XP의 3.0GHz 펜티엄 IV 시스템에서 실행되었다.

4.1 R과 질의의 수가 미치는 영향

그림 5는 $W=50, |O|=50,000$ 로 설정하고, |Q|의 크기를 8,000부터 32,000개까지, R의 크기를 512에서 9,192까지 변화를 주면서 성능을 평가하였다. PAB-기반의 질의색인의 저장비용은 R의 크기가 증가할수록 완만한 증가를 보였다. 이와 대조적으로, CES-기반의 질의색인은 질의 ID를 2차원의 그리드형태로 유지하기 때문에 $R = 8,192$ 일 때 색인저장비용이 700Mbyte정도로 크다.

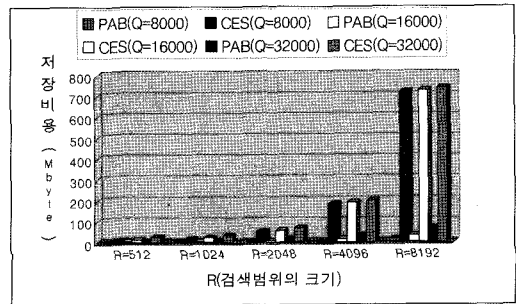


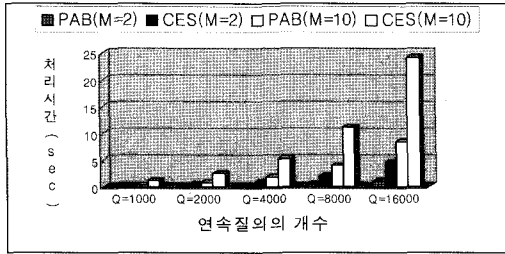
그림 5 질의의 수와 R에 따른 저장비용

4.2 질의의 수와 객체의 수에 따른 영향

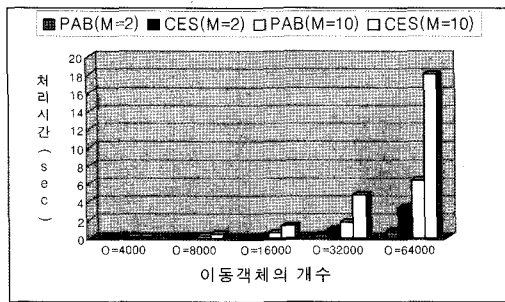
그림 6(a)는 처리시간에 대한 연속질의의 수 |Q|의 영향을 보여준다. 이 실험을 위해, $W=50, L=16$ 및 $|O|=50,000$ 로 설정하였다. |Q|는 1,000부터 16,000까지 변화하고 균일하게 배치하고, $M=2$ 와 $M=10$ 을 사용하였다. PAB-기반의 질의색인을 위한 처리시간은 |Q|가 증가에 따라 점차적인 증가하였다. 그러나 CES-기반의 질의색인은 질의 처리 성능이 크게 저하되었다. 또한 M가 클수록 CES-기반의 질의색인을 위한 처리시간이 더 많이 소요되었다. PAB-기반의 질의색인은 M의 크기에 변화에 큰 효과가 없다. 그림 6(b)는 |O|의 영향을 보여준다. 이 실험을 위해, $W=50, L=16, |Q|=8,000$ 로 설정하고 |O|는 4,000부터 64,000까지 변경한다. 이동객체의 수의 변화에 PAB-기반의 질의색인은 비트연산을 통하여 산출하기 때문에 CES-기반의 질의색인보다 성능이 향상되었다.

5. 결론

본 논문은 연속 범위 질의를 효율적으로 처리하는 새로운 PAB-기반 질의색인기법을 제안하였고 효율적인 질의의 처리를 위하여 점진적인 갱신을 제공한다. 성능평가의 결과는 PAB-기반 질의색인이 기존방법들보다



(a) 질의의 개수에 따른 처리시간



(b) 객체의 개수에 따른 처리시간

그림 6 객체의 수와 질의의 수에 따른 질의처리시간

질의처리시간이나 저장비용에서 탁월한 성능을 보여주었다. 더 중요한 것은 PAB-기반 질의색인이 고차원을 위한 대용량 데이터의 처리에 매우 효율적임을 확인하였다. 향후 연구방향으로 복잡한 속성을 갖고 있는 동적인 연속 범위 질의를 메모리 기반 색인으로 설계할 것이다.

참 고 문 헌

[1] S. Prabhakar, Y. Xia, D. V. Kalashnikov, W. G. Aref, and S. E. Hambrusch, Query Indexing and Velocity Constrained Indexing: Scalable Techniques for Continuous Queries on Moving Objects. In IEEE Trans. Computers, Vol.51, No.10, pp. 1124-1140, Oct. 2002.

[2] K.-L.Wu, S.-K Chen, and P.S. Yu, Efficient Processing of Continual Range Queries for Location-Aware Mobile Services. In Information Systems Frontiers, Vol.7, Nos.4-5, pp. 435-448, Dec. 2005.

[3] K.-L.Wu, S.-K Chen, and P.S. Yu, Processing continual range queries over moving objects using VCR-based query indexes. In Proceedings of IEEE Int'l Conf on Mobile and Ubiquitous Systems: Networking and Services, pp. 226-235, Aug. 2004.

[4] K.-L. Wu, S.-K Chen, and P.S. Yu, Shingle-Based Query Indexing for Location-Based Mobile E-Commerce. In Proceedings of IEEE Int'l Conf.

E-Commerce, July. pp. 16-23, 2004.

[5] D. V. Kalashnikov, S. Prabhakar, W. G. Aref, and S. E. Hambrusch, Efficient Evaluation of Continuous Range Queries on Moving Objects. In Proceedings of Int'l Conf. Database and Expert Systems Applications, Vol. 2453/2002, pp. 25-64, 2002.

[6] B. Gedik, K.-L. Wu, P.S. Yu, and L. Liu, Processing Moving Queries over Moving Objects using Motion-Adaptive Indexes. In IEEE Trans. Knowledge and Data Eng, Vol.18, pp. 651-668, May 2006.

[7] B. Gedik and L. Liu, MobiEyes: Distributed Processing of Continuously Moving Queries on Moving Objects in a Mobile System. In Proceedings of Int'l Conf. Extending Database Technology, 2004.

[8] M.F. Mokbel, X. Xiong, and W. G. Aref, SINA: Scalable Incremental Processing of Continuous Queries in Spatio-Temporal Databases. In Proceedings of ACM SIGMOD Int'l Conf. Management of Data, 2004.

[9] K.-L. Wu, S.-K Chen, and P.S. Yu, Incremental Processing of Continual Range Queries over Moving Objects. In IEEE Trans. Knowledge and Data Eng, Vol.18, pp. 1560-1575, Nov. 2006.

[10] A. Guttman, R-Trees: A Dynamic Index Structure for Spatial Searching. In Proceedings of ACM SIGMOD Int'l Conf, Management of Data, pp. 47-57, 1984.