

# 무선 방송 환경에서 제한된 지역의 NN 질의 처리 방법\*

이명수, 류병걸, 오재오, 이상근  
고려대학교 정보통신대학 컴퓨터통신공학부  
e-mail: {lms9711, smart123, doublefive, yalphy}@korea.ac.kr

## Constrained Nearest Neighbor Query Processing in Wireless Broadcast Environments

Myong-Soo Lee, Byung-Gul Ryu, Jae-Oh Oh, SangKeun Lee  
Division of Computer and Communication Engineering, Korea University

### 요 약

위치기반 서비스는 모바일 기기와 무선 통신 기술의 발달로 인해 유비쿼터스 정보 접근의 요구에 따라 많은 관심을 받고 있다. 위치기반서비스 중에서 제한된 지역의 NN 질의는 무선 통신을 통해 그 수요가 급격히 증가하고 있는 질의 중 하나이다. 무선 방송 채널의 효율적 사용과 제한된 자원을 가진 무선 기기에서 제한된 지역의 NN 질의를 효율적으로 수행하기 위해 무선 방송환경에서 적합한 질의 처리 방법을 제안한다. 본 논문에서는 기존의 무선방송환경에서 적합한 기법인 분산 공간 색인 기법 기반한 제한된 지역의 NN 질의 방법을 제안하였으며 본 기법은 다른 기법에 비해 접근 시간과 튜닝 시간에 대하여 좋은 성능을 보인다.

### 1. 서론

데이터 브로드캐스팅은 사용자에서 언제 어디서나 요구된 정보를 상황 인지를 통해 제공되는 유비쿼터스 컴퓨팅의 중요한 연구 주제중 하나이다 [5][8][9]. 위치 정보의 상황인지는 사용자의 이동성을 고려하여 위치에 기반한 데이터를 제공하는 것이다. 무선 데이터 브로드캐스팅은 많은 사용자에게 동시에 지원할 수 있는 확장성과 부족한 통신 채널을 효율적으로 사용할 수 있다는 점에서 큰 장점을 가지고 있다.

위치기반서비스에서의 중요한 두 개의 질의는 윈도우 질의(Window Query)와  $k$ 개의 최근접점 질의( $k$  Nearest Neighbor Query,  $kNN$ )이다 [4][6][10]. 윈도우 질의는 일정한 형태의 범위 내에 존재하는 객체를 찾는 질의이다.  $kNN$ 질의는 질의점에서 가장 가까운  $k$ 개의 객체를 찾는 질의이다. 본 논문에서는  $kNN$ 의 다른 형태인 특정한 지역내에서의 질의점과 가장 가까운 NN을 찾는 질의인 제한된 지역의 NN 질의 (Constrained Nearest Neighbor, CNN)를 다룬다. “현재의 위치에서 가장 가까운 성북구 내의 중국 음식점을 찾아라”, “현재의 위치의 남쪽에서 가장 가까운 주유소를 찾아라”와 같은 질의는 CNN 질의의 한 예이다 [2].

본 논문에서는 제한된 지역의 NN 질의를 무선 방송환경에서 적용한 기법을 제안한다. 기존의 위치기반서비스의 무선 방송환경에서의 질의 기법은 CNN 질의를 처리하기

에는 접근시간과 튜닝시간에서 비효율적이다. 본 논문은 전체 지도를 힐버트커브 순서에 따라 특징지어진 공간으로 나뉘어진 전체 공간을 색인한 기법인 비트맵 기반의 색인 기법(Bitmap-based Spatial Indexing, BSD)[11]을 이용하여 에너지 효율적인 CNN 질의 기법을 제안한다.

본 논문의 내용은 다음과 같다. 2장에서는 관련 연구를 알아보고, 3장에서는 CNN 질의를 처리하기 위한 두 기본 알고리즘을 제안한다. 4장에서는 실험 결과를 보여주고 5장에서 결론을 내린다.

### 2. 관련 연구

#### 2.1 제한된 지역에서의 NN 질의

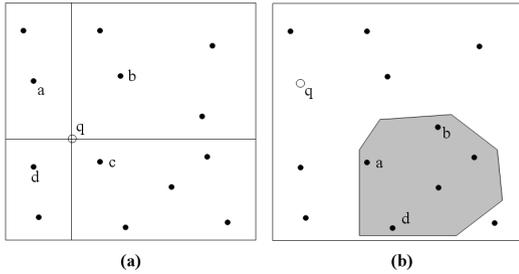
제한된 지역에서의 NN을 찾기 위한 질의는 기존의 디스크 기반의 공간 데이터베이스에서는 많은 연구가 있어 왔다. R-tree[4]를 기반으로 CNN과 윈도우 질의를 처리하기 위한 새로운 기법이 [2]에서 제안되었다. CNN 질의는 특징에 따라 다음과 같은 3가지로 분류할 수 있다.

- (1) 질의점에서 이동방향을 제약사항으로 가지는 경우 : 지금 위치에서 특정 방향을 제약사항을 가지면서 가장 가까운 NN을 찾는 경우이다. 이때, 특정 방향은 남쪽, 남동쪽과 같이 표현할 수 있다. 그림1(a)는 이동방향을 제약사항으로 가지는 예제이다. “지금 위치에서 북동쪽으로 가장 가까운 도시를 찾아라”라는 질의에서 결과값은 “b”가 된다.
- (2) 질의점에서 특정한 지역을 제약사항으로 가지는 경

\* 이 연구에 참여한 연구자는 ‘2단계 BK21’의 지원비를 받았음

우 : 특정 지역내의 객체에 대해서 지금 위치와 가장 가까운 최근접을 찾는 경우이다. 이때, 특정 지역은 안암동, 서울시, 고려대학교와 같이 표현하며, 그 지역은 다각형으로 표현할 수 있다. 그림1(b)는 특정한 지역을 제약사항으로 가지는 예제이다. “지금 위치에서 가장 가까운 서울 학교를 찾아라”라는 질의에서 결과값은 “a”가 된다.

(3) 질의점에서 도로를 제약사항으로 가지는 경우 : 특정 지역에서 이동시 도로로만 이동한다고 가정할 경우 가장 가까운 최근접을 찾는 경우이다. 각 도로는 꼭지점과 간선을 가지는 그래프로 표현할 수 있다.



(그림 1) CNN의 예제

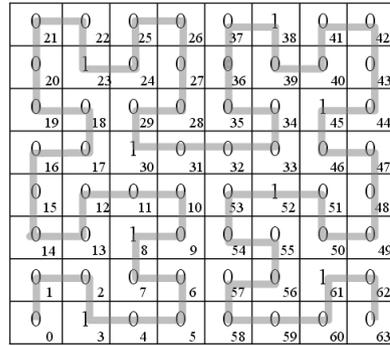
2.2 비트맵 기반의 공간 색인

우리의 이전의 연구[11]에서는 무선 방송 환경에서 위치 기반서비스의 색인으로 가장 좋은 성능을 보인 분산공간 색인 기법(Distributed Spatial Index, DSI)를 기반으로 한 비트맵 기반의 공간 색인(Bitmap-based Spatial Index, BSI)을 제안하였다. 이 기법은 비트맵 기반 공간 색인 구조의 설계를 위하여 다음을 고려한다. 비트맵 기반 공간 색인 구조는 각 힐버트 커브 값[3]의 데이터 유무를 나타내는 맵 정보를 갖는다. 사용자가 필요한 데이터만을 읽음으로써 질의 수행 성능을 향상시킬 수 있다. 비트맵 기반의 색인은 기존 분산공간색인 구조에 추가되어 전체 데이터를 각 세부공간별로 분할하고 각 세부공간은 색인 테이블과 비트맵을 갖는다. 비트맵 공간 색인 구조는 모든 힐버트 커브값의 유무를 대표한다 [11].

그림 2는 비트맵기반의 공간 색인의 핵심인 비트맵의 구조의 한 예제이다. 이 예제에서는 힐버트 순서가 3이고, 모든 데이터의 유무는 64비트들로 나타낸다. 각각의 격자는 하나의 비트값, 즉 1과 0으로 표현된다. 격자 내에 데이터가 존재할 경우에는 1, 존재하지 않는 경우는 0으로 표현하며 이를 힐버트 커브 순서대로 나열하여 하나의 값으로 표현한다. 결과적으로 64비트의 값을 통해서 전체 맵 정보를 알 수 있다.

3. 무선 방송 환경에서의 제한된 지역의 NN 질의 기법

3.1 기본적 기법(N\_CNN)

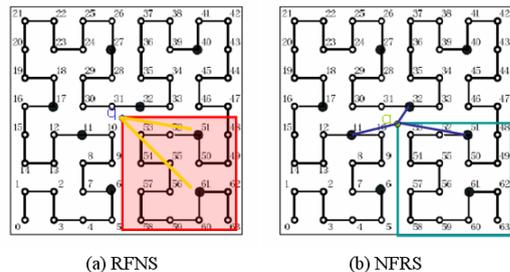


BSI  
0010000100000000...0100<sub>(2)</sub>

(그림 2) 비트맵 기반의 공간 색인 [11]

무선 방송 환경에서 제한된 지역의 NN 질의 기법을 수행하기 위해서 우리는 분산공간색인(DSI)[7]를 이용한 기본 기법(Naive CNN, N\_CNN)을 제안한다. 본 기본 기법은 두가지 다른 형태의 기법으로 나눌 수 있는 있다. 첫 번째 기법은 NN이 될 후보를 추출한 후 제한된 지역 안에서 후보의 포함 여부를 확인한다. 방송 채널에서 DSI를 사용함으로써 사용자는 첫 번째 방송주기에서 제한된 지역을 탐색하고 질의점과 모든 제한된 지역내의 데이터와의 거리를 계산한다. 제한된 지역의 크기와 위치에 따라 범위 질의를 먼저 수행하고 NN 질의를 하는 것이 더 효율적이다. 이 기법을 RFNS (Range-First NN-Second)이라 명명한다.

두 번째 기본 기법은 다음과 같은 단계를 거친다. 우선 제한되지 않은 지역의 모든 NN들을 계산하여 이 후보들을 질의점과의 거리 순으로 나열한 후 각각의 후보들이 제한된 지역에 포함되는지 여부를 확인한다. 사용자가 제한된 지역에 포함되는 후보가 나타날때까지 모든 데이터를 검색해 나간다. 이 경우 질의점 근처에 데이터가 모두 제한된 지역에 포함되지 않을 경우 많은 후보가 발생하고 많은 데이터를 검색해야 한다는 문제점이 발생한다. 이 기법을 NFRS (NN-First Range-Second)이라 명명한다.



(a) RFNS (b) NFRS

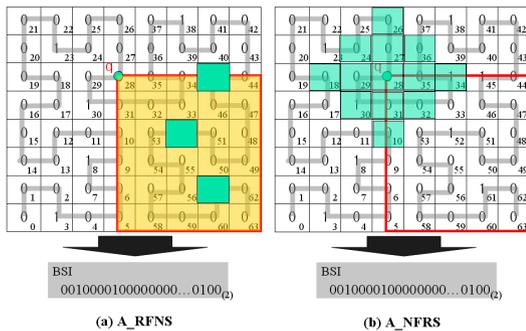
(그림 3) 기본적 기법(N\_CNN)

### 3.2 개선된 기법(A\_CNN)

기본적 기법에서 사용한 DSI는 모든 맵 정보를 알기 위해서는 접근 시간이 길어지는 단점이 있다. BSI는 맵 정보를 한번에 알 수 있는 장점이 있어 DSI의 단점을 보완할 수 있다. 또한 이 기법은 보다 적은 수의 패킷만을 읽음으로써 튜닝타임을 줄여 에너지 소모량을 감소시킬 수 있다. 우리는 BSI 색인에 알고리즘을 추가하여 CNN 알고리즘을 제안한다. 기본적 기법(N\_CNN)과 비슷하게 여기에서도 두 가지 알고리즘을 제안한다.

첫 번째 기법은 A\_RFNS (Advanced Range-First NN-Second)이며 이는 기본적 기법과 비슷하게 두 단계를 거친다. 이 방법에서 먼저 무선 방송 채널에서 BSI 색인 정보를 먼저 읽는다. BSI를 읽은 후에 우리는 제한된 지역 내에서 HC 값과 BSI를 비교한다. 질의의 힐버트 커브 값에 대응하는 BSI의 비트 값이 1이면, 이 점은 위치가 검증된 데이터 객체이고 이를 실제 데이터를 탐색하는 후보군에 포함한다. 다음 단계로, 후보군에 있는 모든 데이터와 질의점과의 거리를 측정한다. 마지막으로, 사용자는 NN을 알게 되고 이를 바탕으로 실제 데이터를 검색하게 된다.

두 번째 기법은 A\_NFRS (Advanced NN-First Range-Second)이며 이 기법은 A\_RFNS와 마찬가지로 다음과 같은 과정을 거친다. 우선 사용자는 무선방송 채널로부터 BSI 정보를 읽고 이 정보를 통해 주어진 질의점의 거리 순서에 따라 후보점들이 BSI의 비트 정보를 알아낸다. 다음으로 각 데이터가 제한된 지역 내에 포함이 되는지 안되는지 검색한다. 데이터가 제한된 지역 내에 포함되면 데이터를 결과셋에 포함시키고 그렇지 않으면 비트값에 따라 계속해서 계산해 나간다. 계산은 k개의 데이터 객체 결과값이 나올때까지 계속되며 마지막으로 주어진 색인을 가지고 실제 데이터를 검색한다.



(그림 4) 개선된 기법(A\_CNN)

### 4. 성능 평가

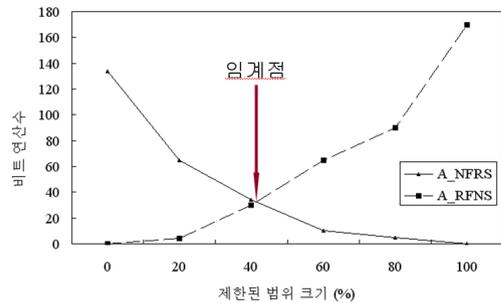
제안한 기법에 대한 성능을 평가하기 위해서 다음과 같은 두 가지 측정 요소를 사용한다. 방송기법에서의 색인

기술의 주요 성능 측정 요소로는 결과값을 얻는 시간인 접근시간(Access Time)과 사용자가 활동 모드로 전력을 소모하게 되는 시간인 튜닝 시간(Tuning Time)이므로 이 두 가지 요소를 측정하였다. 본 논문에서는 튜닝시간은 모바일 기기가 전력을 소모하는 방송채널에서 패킷을 읽어들이는 시간을 의미한다. 데이터 집합은 5922개의 객체를 가지고 있으며, R-tree 포탈[1] 웹사이트에 있는 그리스의 도시와 마을의 위치인 데이터를 사용하였다. 기본적 버킷 크기는 64bytes이고 kNN의 기본 k값은 10이다.

### 4.1 임계점 계산

본 논문에서는 제안한 개선된 기법(A\_CNN)은 두 방법으로 탐색한다. 하나는 범위를 먼저 탐색하고 NN을 찾는 방법(A\_RFNS)이고 다른 방법은 NN을 찾고 범위를 탐색하는 방법(A\_NFRS)이다. 두 방법의 트레이드오프(tradeoff)는 제한된 지역 R이 전체 데이터 공간에 비해 크면 질의점에서 R의 모든 점들까지의 거리를 모두 계산해야 해서 계산 비용이 많이 든다. 즉 이 경우 A\_RFNS가 A\_NFRS보다 더 효율적이다. 그러나 제한된 지역 R이 작은 수의 데이터만을 포함하고 있을 경우 많은 데이터를 거리 순서대로 배열하여 많은 데이터를 검색해야 하므로 A\_NFRS가 A\_RFNS보다 효율적이다.

따라서, 우리는 제한된 범위의 크기에 따라 임계점을 설정하여 각 상황에 맞는 방법을 사용함으로써 좋은 성능을 보일 수 있다. 그림 5는 어떻게 임계점을 설정하는지 보여준다. BSI에서 두 기법은 비트 계산의 수에 따라 성능이 달라진다. 비트 계산은 모바일 클라이언트의 에너지 효율과 관련된다. 이 그림에서 범위 크기의 비율이 40%보다 작으면 A\_NFRS가 A\_RFNS보다 효율적이고 범위 크기의 비율이 40%보다 높으면 그 반대의 성능을 보인다. 따라서 우리는 질의의 제한된 지역의 범위크기에 따라서 기법을 선택할 수 있다. 이러한 임계점은 기본적 기법(N\_CNN)에도 동등하게 적용된다.

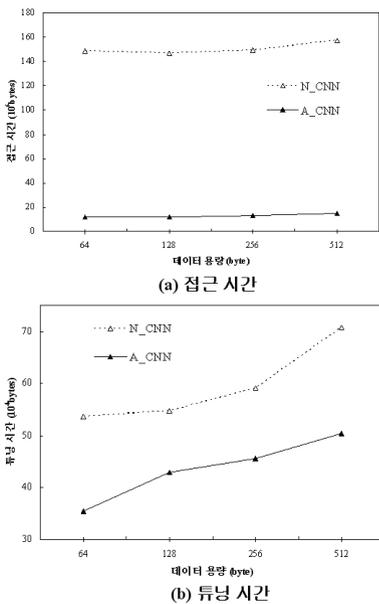


(그림 5) 임계점 계산

### 4.2 접근 시간과 튜닝 시간

그림 6는 기본적 기법(N\_CNN)과 개선된 기법(A\_CNN)

의 접근 시간과 튜닝 시간을 평가한다. 각 성능은 데이터 패킷의 크기를 64byte에서 512byte까지 다양한 환경을 주어 평가한다. kNN의 k개는 10개로 설정했으며 실제 데이터를 사용했다. 그림 6(a)에서 보는 바와 같이 A\_CNN은 N\_CNN에 비해 접근 시간은 평균적으로 10배 이상 좋아짐을 보인다. 이는 BSI를 이용한 A\_CNN이 모든 데이터를 검색하지 않고 단지 BSI만을 확인하여 모든 데이터의 위치 유무를 파악함으로써 한번의 방송 주기내에 질의를 처리함으로써 접근 시간을 크게 줄일 수 있다. 그림 6(b)에서는 A\_CNN이 N\_CNN에 비해 튜닝 시간을 평균적으로 2배 이상 좋아짐을 보인다. 이는 개선된 기법이 기본적 기법보다 에너지 활용 측면에서 높은 효율을 보임을 나타낸다. 이는 개선된 방법이 실제 데이터를 읽지 않고도 데이터의 위치를 파악 가능하여 많은 데이터를 읽지 않고 생략이 가능함으로써 튜닝시간이 줄어 에너지 소모량을 줄인다.



(그림 6) 접근 시간과 튜닝시간

## 5. 결론 및 향후 연구

본 논문은 무선 방송 환경에서 제한된 지역에서의 NN 질의(CNN)를 제안하였다. 우리는 기본적 기법(N\_CNN)과 개선적 기법(A\_CNN) 두 방법을 제안하였고 각 기법의 내부 두 가지 방법을 임계점을 이용해 좋은 성능을 보이는 새로운 기법을 제안하였다. 실험 결과 비트맵 기반의 공간 색인(BSI)를 이용한 A\_CNN이 접근 시간과 튜닝 시간에서 N\_CNN에 비해 좋은 성능을 보였다. 이는 A\_CNN이 실제 모바일 기기에서 사용시 에너지와 질의의 신속성 측면에서 장점이 있다는 것을 의미한다.

향후 연구로는 BSI 색인 크기를 줄여 성능을 향상시킬 것이며 무선 방송 환경에서의 새로운 형태의 질의를 처리하는 시스템을 개발할 것이다.

## 참고문헌

- [1] Spatial datasets. <http://www.rtreportal.org/spatial.html/>.
- [2] H. Ferhatosmanoglu, I.Stanoi, D.Agrawal, and A. Abbadi. Constrained Nearest Neighbor Queries. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, pages:257-276, 2001.
- [3] C. Gosman and M. Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, pages 794-797, 1996.
- [4] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 47-54, 1984.
- [5] T. Imielinski, S. Viswanathan, and B. R. Badrinath. Data on air organization and access. *IEEE Transactions on Knowledge and Data Engineering*, 9(3), 1997.
- [6] D. L. Lee, W. C. Lee, J. Xu, and B. Zheng. Data management in location-dependent information services: Challenges and issues. *IEEE Pervasive Computing*, 1(3):65-72, 2002.
- [7] W. C. Lee and B. Zheng. DSI: A fully distributed spatial index for location-based wireless broadcast services. In *Proceedings of the International Conference on Distributed Computing Systems*, pages 349-358, 2005.
- [8] B. Zheng and D. L. Lee. Information dissemination via wireless broadcast. *Communications of the ACM*, 48(5):105-110, 2005.
- [9] B. Zheng, W. C. Lee, and D. L. Lee. Spatial index on air. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, pages 297-304, 2003.
- [10] B. Zheng, W. C. Lee, and D. L. Lee. Spatial queries in wireless broadcast systems. *Wireless Network*, 10(6):723-736, 2004.
- [11] 신하늘, 이명수, 최재호, 이상근. 무선방송환경에서의 비트맵 기반 공간색인 기법. *데이터베이스연구회지*, 23(1):23-36, 2007