

무선방송환경에서 에너지 효율적인 제한된 최근접 질의 처리

이 명 수[†] · 류 병 곁^{††} · 오 재 오^{††} · 이 상 근^{†††}

요 약

위치기반 서비스는 무선기기와 무선 통신 기술의 발달로 인해 유비쿼터스 정보 접근의 요구에 따라 많은 관심을 받고 있다. 위치기반 서비스 중에서 제한된 최근접 질의는 무선 통신을 통해 사용자의 요구를 만족시키게 하는 중요한 질의 중 하나이다. 무선 방송 채널의 효율적 사용과 제한된 자원을 가진 무선기기를 효율적으로 사용하기 위해 제한된 최근접 질의를 효율적으로 수행하기 위해 무선방송환경에서 적합한 질의 처리 방법을 제안한다. 우선 지역을 제한조건으로 가지는 최근접 질의 기법을 제안하고 다음으로 다양한 제한조건을 수행할 수 있는 최근접 탐색 기법을 적용하고자 한다. 본 논문에서는 기존의 무선방송환경에서 적합한 기법인 분산공간색인과 비트맵 기반의 공간색인 기법에 기반한 제한된 최근접 질의 방법을 제안한다.

키워드 : 위치기반 서비스, 최근접 탐색(NN), 튜닝 시간, 무선 방송 환경, 제한조건

Energy-efficient Query Processing of Constrained Nearest Neighbor Queries on the Wireless Broadcasting Environments

Myong-Soo Lee[†] · Byung-Gul Ryu^{††} · Oh, Jae-Oh^{††} · SangKeun Lee^{†††}

ABSTRACT

Location based services (LBSs) have drawn huge attention as the growing number mobile devices and wireless technology demand more pervasive information access. In LBSs, Constraint nearest neighbor queries are one of the important queries of human to fulfill his desired quickly through wireless technology. We propose new query processing of constrained nearest neighbor query on the air to support mobile clients which demand optimal uses of wireless broadcast channel and using minimum battery power of client. First we proposed NN query processing with constrained region and then explain the novel NN query processing with various types of constraints. We have proposed novel algorithms to support Constrained Nearest Neighbor queries on the air based on Distributed Spatial Index and Bitmap-based Spatial Index.

Keywords : Location-based Services, Nearest Neighbor(NN), Tuning Time, Wireless Broadcasting, Constraint

1. 서 론

데이터 브로드캐스팅은 사용자에서 언제 어디서나 요구된 정보를 상황 인지를 통해 제공되는 유비쿼터스 컴퓨팅의 중요한 연구 주제 중 하나이다 [7, 10, 11]. 위치 정보의 상황인지는 사용자의 이동성을 고려하여 위치에 기반한 데이터를 제공하는 것이다. 무선 데이터 브로드캐스팅은 많은 사용자에게 동시에 서비스를 지원할 수 있는 확장성과 부족한 통신 채널을 효율적으로 사용할 수 있다는 점에서 큰 장점을

가지고 있다.

위치기반 서비스에서의 중요한 두 개의 질의는 윈도우 질의(Window Query)와 k개의 최근접 질의(k Nearest Neighbor Query, kNN)이다 [6, 8, 12]. 윈도우 질의는 일정한 형태의 범위 내에 존재하는 객체를 찾는 질의이다. kNN 질의는 질의점에서 가장 가까운 k개의 객체를 찾는 질의이다. 본 논문에서는 kNN의 다른 형태인 특정한 제한조건을 가진 질의 점과 가장 가까운 최근접을 찾는 질의인 제한된 최근접 질의(Constrained Nearest Neighbor, CNN)를 다룬다. “현재의 위치에서 가장 가까운 자장면 값이 3000원 이하인 중국 음식점을 찾아라”, “현재의 위치의 남쪽에서 가장 가까운 주유소를 찾아라”와 같은 질의는 CNN 질의의 예들이다 [4].

본 논문에서는 제한된 최근접 질의를 무선 방송환경에서 적용한 기법을 제안한다. 우선 제한된 지역의 최근접 질의 기법을 제안한 우리의 이전의 연구[2]와 마찬가지로 지역을 제한조건으로 가진 N_CNN과 A_CNN 기법을 제안한다. 다

※ 이 연구에 참여한 연구자는 '2단계 BK21사업'의 지원비를 받았음.
※ 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-313-D00653).
† 준 회 원 : 고려대학교 컴퓨터학과 박사과정
†† 준 회 원 : 고려대학교 컴퓨터학과 석사과정
††† 정 회 원 : 고려대학교 컴퓨터학과 부교수
논문접수: 2008년 11월 7일
수정일: 1차 2008년 12월 16일
심사완료: 2009년 2월 18일

음으로 지역 뿐만 아니라 다양한 제한조건을 가진 경우의 최근접 질의처리기법으로서 AS_CNN과 AA_CNN 기법을 적용한다. 기존의 위치기반서비스의 무선방송환경에서의 질의 기법은 CNN 질의를 처리하기에는 비효율적이다. 기존의 기법들은 제한사항의 조건에 맞는지 확인하기 위해 많은 실제 데이터를 탐색하기 때문에 많은 접근시간과 튜닝시간을 필요로 한다. 길어진 접근시간과 튜닝시간은 질의의 수행 시간과 에너지 효율에서 나쁜 성능을 나타낸다. 본 논문은 전체 지도를 힐버트 커브 순서에 따라 특정된 공간으로 나누어진 전체 공간을 색인한 기법인 비트맵 기반의 색인 기법(Bitmap-based Spatial Indexing, BSI) [3]을 이용하여 에너지 효율적인 CNN 질의 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 알아보며, 3장에서는 제한된 지역에서의 최근접 질의를 처리하기 위한 두 기본 알고리즘을 제안한다. 4장에서는 다양한 제한조건을 가진 최근접질을 처리하기 위한 두 알고리즘을 제안한다. 5장에서는 실험 결과를 보여주고 6장에서 결론을 내린다.

2. 관련 연구

2.1 제한 조건으로 제한지역을 가진 경우의 최근접 질의

제한된 지역에서의 최근접을 찾기 위한 질의는 기존의 디스크 기반의 공간 데이터베이스에서는 많은 연구가 있어 왔다. R-tree[6]를 기반으로 CNN과 윈도우 질의를 처리하기 위한 새로운 기법 [4]에서 제안되었다. CNN 질의는 특징에 따라 다음과 같은 3가지로 분류할 수 있다.

- (1) 질의점에서 이동방향을 제한사항으로 가지는 경우 : 지금 위치에서 특정 방향을 제한사항을 가지면서 가장 가까운 최근접을 찾는 경우이다. 이때, 특정 방향은 남쪽, 남동쪽과 같이 표현할 수 있다. (그림 1(a))는 이동방향을 제한사항으로 가지는 예제이다. “지금 위치에서 북동쪽으로 가장 가까운 도시를 찾아라”라는 질의에서 결과값은 “b”가 된다.
- (2) 질의점에서 특정한 지역을 제한사항으로 가지는 경우 : 특정 지역내의 객체에 대해서 지금 위치와 가장 가까운 최

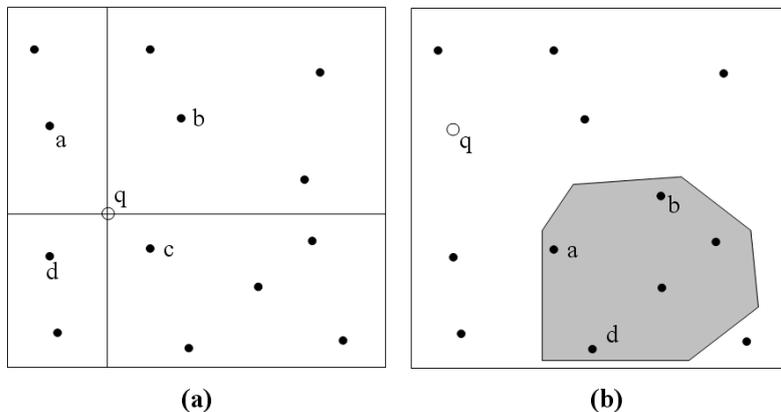
근접을 찾는 경우이다. 이때, 특정 지역은 안암동, 서울시, 고려대학교와 같이 표현하며, 그 지역은 다각형으로 표현할 수 있다. (그림 1(b))는 특정한 지역을 제한사항으로 가지는 예제이다. “지금 위치에서 가장 가까운 서울 학교를 찾아라”라는 질의에서 결과값은 “a”가 된다.

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

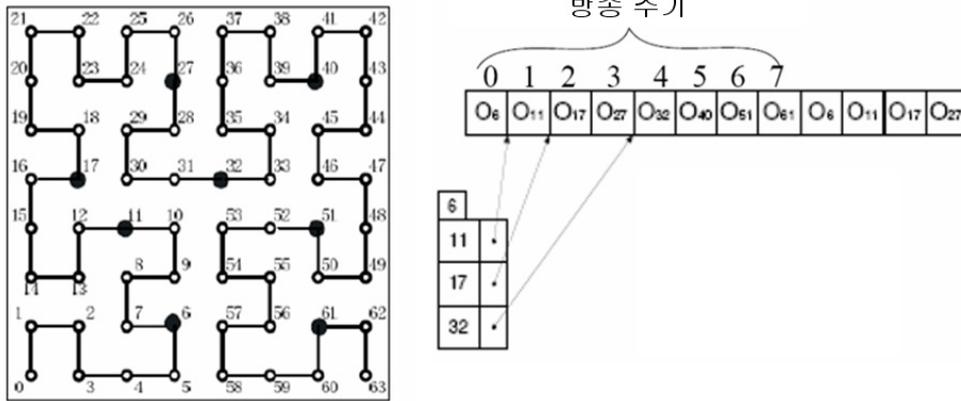
2.2 분산공간색인

서버 클라이언트 모델에서는 위치기반 서비스를 지원하기 위해 R-tree [6]등 많은 색인이 개발되어왔다. 하지만, 무선 방송 환경에서는 기존의 위치기반 서비스의 색인을 사용하기에는 접근 시간과 튜닝 시간이 늘어나는 단점이 있다. 무선방송환경의 순차적인 특징을 적용하기 위해 몇 개의 색인이 제안되었다. 제안된 많은 색인 중에 최근접 질의와 윈도우 질의를 모두 수행하면서 접근 시간과 튜닝 시간이 최적화된 색인이 분산공간색인 기법 (Distributed Spatial Index, DSI) [9]이다.

분산공간색인은 가까운 거리에 위치하는 데이터를 방송 채널에서 가깝게 스케줄링하는 스페이스 필링 커브(Space-Filling Curve) [5] 중 하나인 힐버트 커브(Hilbert Curve)를 기반으로 한다. 힐버트 커브는 무선 방송 채널에서 상에서 다차원 공간의 데이터를 근접한 위치에 위치시킨다. 이때, 커브의 차원은 공간에서 객체의 분할 정도에 따라 결정되며, 객체들의 밀집도가 높을수록 높은 차원을 필요하게 된다. 분산공간색인은 잦은 접속단절이나, 통신에러가 발생하는 무선 방송 환경을 고려한 기법으로 무선 방송 환경에 특화되어 적용할 수 있는 선형구조이며, 각 데이터간 링크를 서로 공유하는 다양한 탐색 경로를 가진다. 분산공간색인은 방송 주기 내에서 충분히 분산된 색인 구조를 가지고 있어서 질의 수행시 짧은 접근 시간을 가진다. 이러한 특성 때문에 에러 경향이 있는 무선 방송 환경에 성능이 뛰어나다. (그림 2)는 분산공간색인의 한 예를 보여준다. 왼쪽의 그림에서 한 방송주기



(그림 1) CNN의 예제

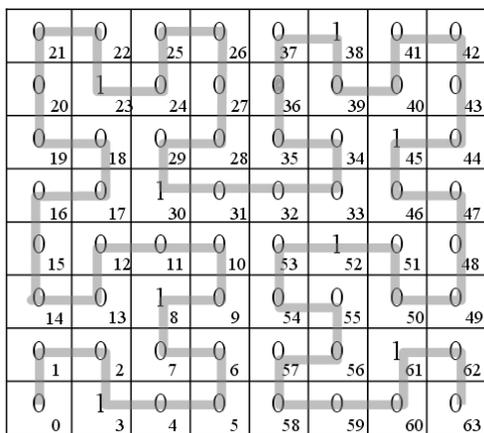


(그림 2) 분산공간색인의 예제 [9]

에서 0번부터 7번까지의 각 버킷은 각각 하나의 색인을 가지고 있다. (그림 2)의 오른쪽 그림의 하단은 6번 버킷의 색인을 보여주고 있는데 이 색인에는 다음 버킷(11번), 두 번째 버킷(17번), 4번째 버킷(32번)에 대한 포인터를 가지고 있다. 모든 버킷에는 6번 버킷과 같은 형태의 색인이 포함되어 있어 전반적으로 다중 색인 구조를 가지고 있다.

2.3 비트맵 기반의 공간 색인

무선 방송 환경에서 위치기반 서비스의 색인으로 가장 좋은 성능을 보인 분산공간색인 기법(Distributed Spatial Index, DSD)를 기반으로 한 비트맵 기반의 공간 색인(Bitmap-based Spatial Index, BSI) [3]이 제안되었다. 비트맵 기반 공간 색인 구조는 각 힐버트 커브값 [4]의 데이터 유무를 나타내는 맵 정보를 갖는다. 이를 통해 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.



(그림 3) 비트맵 기반의 공간 색인 [3]

(그림 3)은 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.

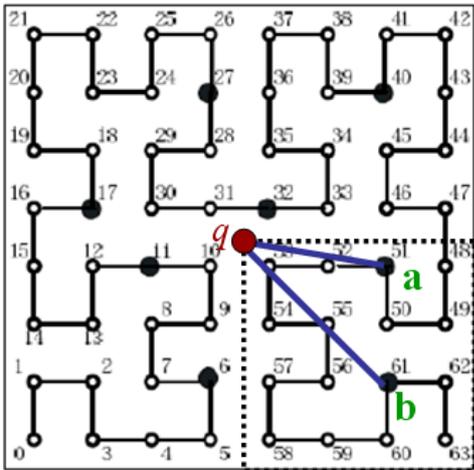
3. 무선 방송 환경에서의 제한된 지역의 최근접 질의 기법

3.1 기본 기법(N_CNN)

무선 방송 환경에서 제한된 지역의 최근접 질의 기법을 수행하기 위해서 우리는 분산공간색인(DSI) [9]를 이용한 기본 기법(Naive CNN, N_CNN)을 제안한다. 본 기법은 해당 지역 내의 최근접이 될 후보를 추출한 후 각 후보들과 질의 점과의 거리를 측정하여 가장 가까운 질의점을 찾는다. 본

입력: 질의점 q , 제한된 지역 w ;
출력: 제한된 지역내의 최근접 객체;
과정:
 01: w 지역내에 포함된 HC(힐버트 커브)값을 계산하여 타겟 집합 H 를 구성;
 02: 최소거리 $R = \infty$, 최근접 객체 $O = \text{null}$;
 03: **While** H 내에 타겟이 남아있음 **Do**
 04: H 내의 HC값이 가장 빠른 타겟을 분산공간색인을 통해 검색;
 05: **If** (타겟내에 객체가 존재)
 06: 타겟내의 객체 O_i 와 질의점과의 거리 R_i 를 계산;
 07: **If** ($R_i < R$)
 08: { $R = R_i$; $O = O_i$; }
 09: **End If**
 10: **End While**
 11: $H = H - [\text{현재의 타겟}]$;
 12: **End While**;
 13: 최근접 객체 반환;

(그림 4) N_CNN 알고리즘



(그림 5) N_CNN의 예제

기법의 알고리즘은 (그림 4)과 같다. 본 기법은 방송 채널에서 분산공간색인을 사용함으로써 제한된 지역내의 객체를 포함하는 타겟 집합 H를 선정한다. 다음으로 H내의 각 타겟의 객체를 검색하여 각 객체 O_i 와 질의점과의 거리를 측정하여 H내의 객체를 다 읽고 난후 가장 가까운 객체를 결정하게 된다. (그림 5)는 본 기법의 예제이다. 본 그림에서 q는 질의 점이고, 점선의 사각형은 제한된 지역이다. 본 기법은 제한된 지역내의 타겟 [48,63]으로부터 분산공간색인을 검색하여 객체 a와 b를 각각 검색하고 이들 객체와 질의점과의 거리를 측정하여 가장 가까운 객체 a를 최종 결정한다.

3.2 개선된 기법(A_CNN)

기본 기법에서 사용한 분산공간색인은 모든 맵 정보를 알기 위해서는 접근 시간이 길어지는 단점이 있다. 비트맵 기반의 공간색인(BSI)은 맵정보를 한 번에 알 수 있는 장점이 있어 분산공간색인의 단점을 보완할 수 있다. 또한 이 기법은 보다 적은 수의 패킷만을 읽음으로서 튜닝타임을 줄여 에너지 소모량을 감소시킬 수 있다. 우리는 BSI 색인에 알고리즘을 추가하여 CNN 알고리즘을 제안한다. 본 기법은 두 가지 세부 알고리즘을 제안한다.

첫 번째 기법은 A_RFNS (Advanced Range-First NN-Second)이다. 본 기법은 제한된 지역 내의 객체를 검색하여 객체별로 질의점과의 거리를 측정하는 방법이다. 본 기법의 알고리즘은 (그림 6)과 같다. 알고리즘에서와 같이 본 기법은 두 단계를 거친다. 이 방법에서 먼저 무선 방송 채널에서 BSI 색인 정보를 먼저 읽는다. BSI를 읽은 후에 우리는 제한된 지역 w내에서 HC 값과 BSI를 비교한다. 질의의 힐버트 커브 값에 대응하는 BSI의 비트 값이 1이면, 이 점은 위치가 검증된 데이터 객체이고 이를 실제 데이터를 탐색하는 후보 집합 H에 포함한다. 다음 단계로, 후보 집합 H에 있는 모든 HC값과 질의점과의 거리를 측정한다. 마지막으로, 사용자는 최근접 객체를 알게 되고 이를 바탕으로 실제 데이터를 검색하게 된다. (그림 8(a))은 본 기법의 예제이다. 그림에서 q는 질의점이고 점선의 사각형은 제한된 지역이다. 본 기법

```

입력: 질의점 q, 제한된 지역 w;
출력: 제한된 지역내의 최근접 객체;
과정:
01: 방송 채널에서 BSI 검색;
02: BSI 값을 통해 w내의 비트가 1인 HC(힐버트 커브)값을 H 집합에 포함
02: 최소거리 R = ∞, HC = null;
03: While H 내에 HC값 남아있음 Do
04:     H내의 HC값중 하나를 선택;
07:     HC값과 질의점과의 거리 Ri를 계산;
06:     If (Ri < R)
07:         { R = Ri; HC = 현재의 HC값; }
08:     End If
09:     H = H - [현재의 HC값];
10: End While;
11: HC를 통해 최근접 객체 반환;
    
```

(그림 6) A_RFNS(A_CNN) 알고리즘

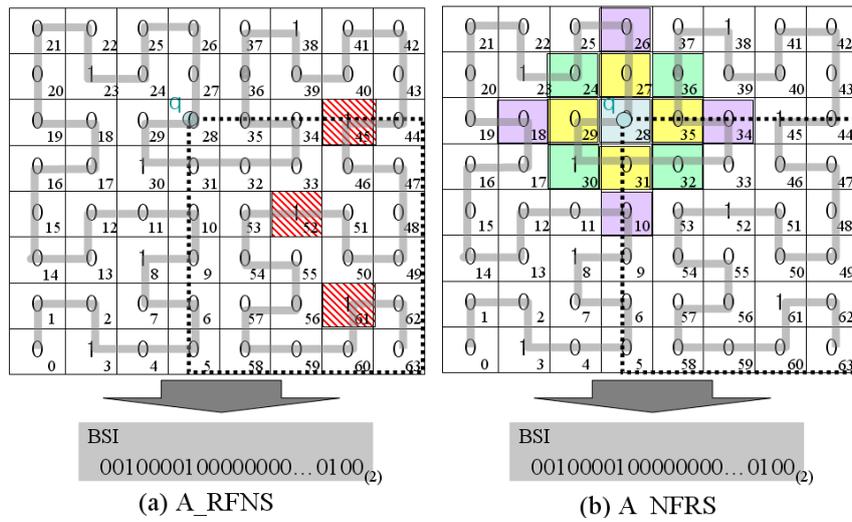
은 방송 채널에서 BSI를 읽고 제한된 지역 내의 1의 비트를 가진 HC값 (45,52,61)과 질의점과의 거리를 유클리디안 거리로 계산하여 그중 가장 가까운 값, 52를 찾아낸다. 최근접 객체의 HC값이 52이므로 이를 분산공간색인을 통해 실제 데이터를 검색하게 된다.

두 번째 기법은 A_NFRS (Advanced NN-First Range-Second)이며 이 기법은 A_RFNS와 마찬가지로 (그림 7)과 같은 알고리즘을 수행하며 그 과정은 다음과 같다. 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 포함 여부를 검색한다. 데이터가 제한된 지역 내에 포함 되면 데이터를 결과셋에 포함시키고 그렇지 않으면 비트값

```

입력: 질의점 q, 제한된 지역 w;
출력: 제한된 지역내의 최근접 객체;
과정:
01: 방송 채널에서 BSI 검색;
02: flag = 0, A =0, HC = null;
03: While (flag == 0) Do
04:     질의점과 거리 A인 HC값의 비트를 검색;
05:     비트가 1인 HC값을 집합 H에 포함;
06:     While (H내의 HC값이 없을 때) Do
07:         H내의 HC값중 하나를 선택;
08:         If (HC값이 w안에 포함됨)
09:             flag = 1;
10:             HC = 현재의 HC값;
11:         End If
12:     End While
13:     If (flag==1)
14:         break;
15:     End If
12:     A = root(square(A)+1);
13: End While
14: HC를 통해 최근접 객체 반환;
    
```

(그림 7) A_NFRS(A_CNN) 알고리즘



(그림 8) A_CNN의 예제

에 따라 계속해서 계산해 나간다. 1의 비트를 가지고 제한된 지역 내에 포함된 데이터가 발견될되면 알고리즘은 종료되며 이로서 사용자는 최근접 객체의 HC값을 알게 되고 이를 바탕으로 실제 데이터를 검색하게 된다. (그림 8(b))는 본 기법의 실행 예제이다. 그림에서의 q는 질의점이며, 점선의 사각형 w는 제한된 지역이다. 본 기법은 BSI를 읽고 질의점과 가까운 HC의 비트를 검색하게 되는데, 그림에서 보는 바와 같이 우선 질의점이 위치한 28번의 비트값을 읽는다. 이 값은 0이므로, 다음으로 가까운 HC값인 (24,30,32,36)의 비트값을 읽는다. 이들 값 중 30번의 값의 비트값이 1이므로 HC값 30의 위치가 제한된 지역 w내에 포함되는지 여부를 검색한다. 30번은 w외부에 존재하므로 30번은 원하는 데이터가 아니다. 그러므로 다음으로 가까운 HC값인 (10,18,26,34)를 검색하게 된다. 이런 식으로 w내에 위치한 비트가 1인 데이터를 검색해나가다 보면 52번의 HC값이 제한된 지역내의 가장 가까운 데이터임을 알 수 있고, 이 HC값으로부터 실제 데이터를 검색해 나간다.

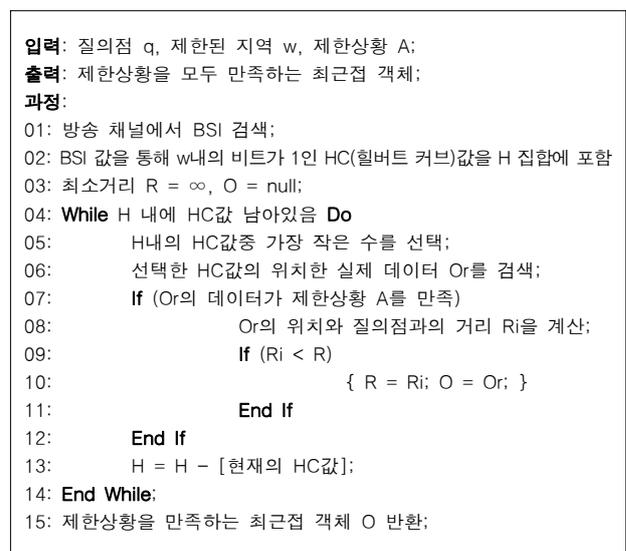
4. 지역 이외의 다양한 제한 상황으로의 확장

지금까지는 제한된 지역에서의 제한상황으로 한정하여 최근접질의 탐색하는 과정을 알아보았다. 하지만, 실제 위치 기반 서비스의 응용에서는 다양한 제한상황이 존재한다. 예를 들면 “휘발유값이 1600 이하이면서 가장 가까운 주유소를 찾아라”, “지금 위치에서 남동쪽으로 가장 가까운 SK 주유소를 찾아라”와 같은 질의이다. 이와 같은 질의는 여러 가지 제한 상황에 때문에 실제 데이터를 검색해서 찾아보아야 할 경우가 많다. 무선 방송환경 상에서의 색인에는 이러한 정보를 충분히 실을 수 없기 때문이다. 본 절에서는 제안한 기법에 본 다양한 제한 상황을 적용시켜 에너지 효율적인 새로운 무선 방송환경에서의 질의 처리 기법을 제안하려 한다. 또한 질의 처리기법은 기본적으로 개선된 기법(A_CNN)

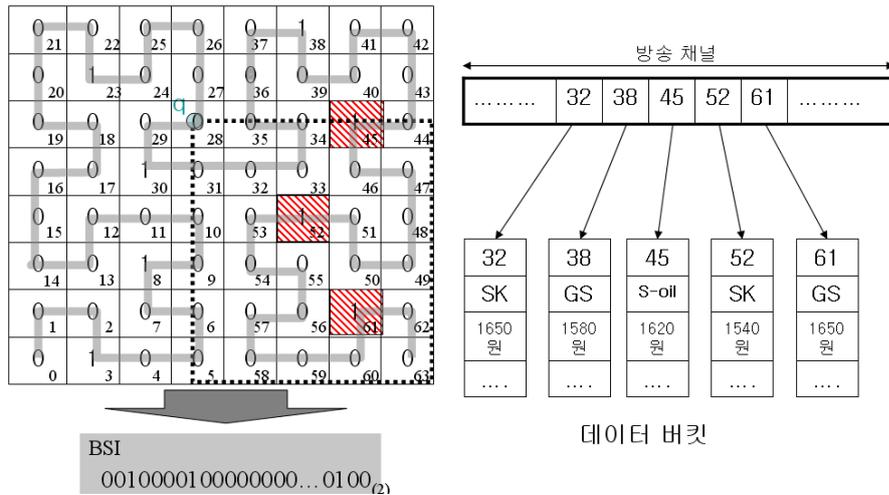
을 바탕으로 질의처리 방법과 실제 데이터 접근 순서를 달리하여 두 가지 방법을 제시한다. 첫 번째는 방송주기 순서대로 데이터를 검색하는 AS_CNN 기법이고 두 번째는 가능성 높은 데이터부터 검색하는 AA_CNN 기법이다.

4.1 AS_CNN

본 기법은 실제 데이터를 방송주기 순서대로 검색하는 AS_CNN(Advanced Sequential Constrained Nearest Neighbor) 기법이다. 우선 CNN 탐색 기법은 앞 절에서 설명한 A_CNN 기법을 기반으로 한다. (그림 9)는 본 기법의 알고리즘을 설명하고 있다. 사용자는 방송채널로부터 우선 비트맵 기반의 공간 색인(BSI)을 읽은 후 이 BSI를 기반으로 제한된 지역내의 모든 후보 최근접 데이터를 찾아낸다. 그 다음으로 후보군 데이터의 방송 순서에 맞추어 하나씩 검색해 나가면서 제한상황에 맞는 데이터를 찾아나간다. 제한상황에 맞는 데



(그림 9) AS_CNN 알고리즘



(그림 10) AS_CNN의 예제

이더를 찾은 경우, 그 데이터의 질의점과의 거리와 그 데이터를 저장하고 계속해서 검색해 나가면서 후보 데이터중에 가장 가까운 데이터를 찾는다.

(그림 10)은 제안한 기법에 대한 한 예제이다. 그림에서 질의는 “질의점 Q에서 남동 방향으로 휘발유값이 1600이하 이면서 가장 가까운 주유소를 찾아라” 이다. 그림에서 보는 바와 같이 방송 채널의 각 데이터 버킷에는 각 주유소의 위치와 주유소의 종류(SK, GS칼텍스, S-oil 등)과 주유소의 휘발류 값 등 각 주유소의 주요 정보가 포함되어 있다. AS_CNN기법을 적용하면 질의점 Q에서 남동쪽인 점선의 사각형 지역 내의 BSI 정보를 분석하여 사각형 지역 내의 1의 비트를 가진 데이터를 검색하여 후보 리스트를 만든다. 그 다음 만든 후보 리스트를 분산공간색인을 이용하여 방송 채널의 순서로 순차적으로 데이터를 읽어 나간다. 그림의 예제에서는 (32, 45, 52, 61)이 후보 데이터가 된다. 이때 방송주기의 순서로 32 → 45 → 52 → 61 순서로 데이터를 검색해 나간다. 우선 32번의 데이터를 읽으면 32번 데이터의 휘발유값이 1650으로 조건에 맞지 않기 때문에 다음 데이터를 읽는다. 다음 데이터인 45번 역시 조건에 맞지 않으므로 다음 데이터인 52번 데이터를 읽게 된다. 52번 데이터는 조건에 맞게 되어 유력한 결과값이 된다. 마지막으로 후보에 남은 데이터인 61번은 계산 결과 52보다 질의점에서 멀어 결국 52번이 이 질의의 답이 되고 질의 수행이 종료된다.

4.2 AA_CNN

본 기법은 실제 데이터를 질의점과 가장 가까운 점부터 찾아나가는 AA_CNN (Advanced Aggressive Constrained Nearest Neighbor)이다. 본 기법은 AS_CNN과 비슷하게 진행되나 실제 데이터를 검색하는 순서가 다르다. 앞의 AS_CNN 기법은 데이터를 순서대로 검색해 나감으로서 접근 시간을 줄일 수 있다는 장점이 있으나 상대적으로 많은 데이터를 읽음으로써 튜닝 시간이 많이 드는 단점이 있다. 이 단점을 보완하기 위해 본 기법은 최근접이 될 가능성이

입력: 질의점 q, 제한된 지역 w, 제한상황 A;
출력: 제한상황을 모두 만족하는 최근접 객체;

과정:

- 01: 방송 채널에서 BSI 검색;
- 02: BSI 값을 통해 w내의 비트가 1인 HC(힐버트 커브)값을 H 집합에 포함
- 02: i=0;
- 03: **While** H 내에 HC값 남아있음 **Do**
- 04: H내의 HC값중 하나를 선택;
- 05: HC값 HC[i]과 질의점과의 거리 R[i]을 계산;
- 06: i=i+1;
- 07: **End While**;
- 08: 배열 R의 값에 대해 오름차순으로 HC를 정렬;
- 09: j=0, O=null;
- 10: **While** (HC[j]!=null) **Do**
- 11: HC[j]의 위치한 실제 데이터 Or를 검색;
- 12: **If** (Or의 데이터가 제한상황 A를 만족)
- 13: O= Or;
- 14: break;
- 15: **End If**
- 16: **End While**;
- 17: 제한상황을 만족하는 최근접 객체 O 반환;

(그림 11) AA_CNN 알고리즘

높은 데이터부터 읽음으로서 튜닝 시간을 줄일 수 있다. (그림 11)은 본 기법의 알고리즘을 보여준다. 본 기법은 BSI를 읽고 난후 제한지역내의 데이터를 거리 순으로 정렬한 후 질의점과의 거리가 가장 가까운 데이터부터 탐색해 나가면서 읽는 데이터 중 제한조건에 해당하는 데이터가 나오면 본 데이터를 반환하면서 질의 탐색이 종료된다.

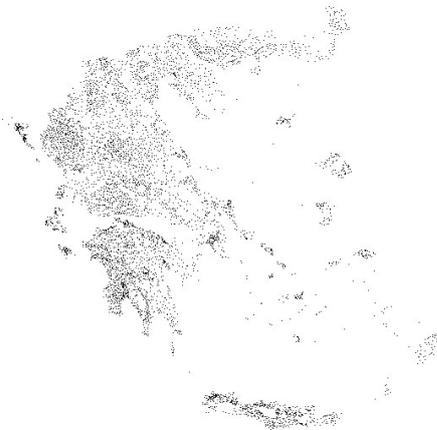
(그림 10)의 예제에서 보면 본 기법은 AS_CNN과 비슷하게 진행되되 실제 데이터를 읽는 순서는 질의점과 가까운 순서인 32 → 52 → 45 → 61 이다. 따라서 우선 32번의 데이터를 읽고 이 데이터의 조건이 질의의 조건과 맞는지 확인한다. 32의 데이터 중 휘발유값은 조건에 맞지 않으므로, 다음 데이터로 진행하게 되고 다음으로 가까운 데이터 52를 검색하게 되고 52번의 데이터는 제한조건에 맞으므로 52번이 조건에 해당하는 가장 가까운 최근접이 되게 된다.

5. 성능 평가

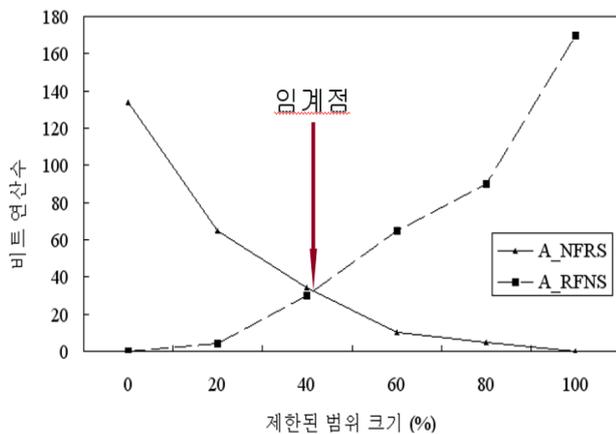
제안한 기법에 대한 성능을 평가하기 위해서 다음과 같은 두 가지 측정 요소를 사용한다. 방송기법에서의 색인 기술의 주요 성능 측정 요소로는 결과값을 얻는 시간인 접근 시간(Access Time)과 사용자가 활동 모드로 전력을 소모하게 되는 시간인 튜닝 시간(Tuning Time)이므로 이 두 가지 요소를 측정하였다. 본 논문에서는 튜닝 시간은 무선기기가 전력을 소모하는 방송채널에서 패킷을 읽어 들이는 시간을 의미한다. 데이터 집합은 5922개의 객체를 가지고 있으며, R-tree 포탈[1] 웹사이트에 있는 그리스의 도시와 마을의 위치인 데이터를 사용하였다 (그림 12 참고). 본 데이터의 위치정보에 실제 데이터를 가공하여 추가하였다. 실제 데이터에는 각 마을 내의 주유소 휘발유값의 최저가를 1.000€에서 1.300€까지 균등하게 분포하여 포함시켰다. 기본적 버킷 크기는 64bytes이고 kNN의 기본 k값은 10이다. 또한, 질의의 제한된 지역의 크기는 전체 맵을 1이라고 가정했을 경우 0.05~0.7으로 설정한다.

5.1 임계점 계산

본 논문에서는 제안한 개선된 기법(A_CNN)은 두 방법으



(그림 12) 실험에 사용된 그리스의 도시와 마을의 위치 [1]



(그림 13) 임계점 계산

로 탐색한다. 하나는 범위를 먼저 탐색하고 최근접을 찾는 방법(A_RFNS)이고 다른 방법은 최근접을 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 돼서 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.

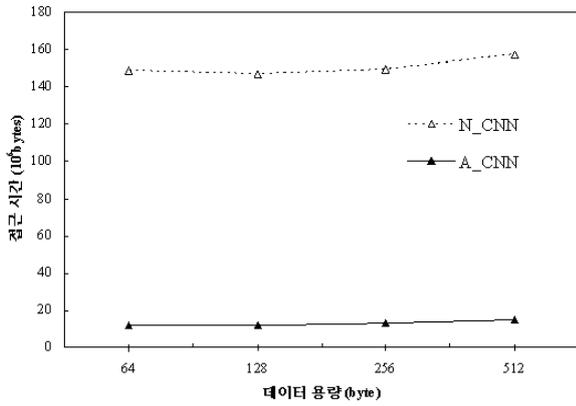
우리는 제한된 범위의 크기에 따라 임계점을 설정하여 각 상황에 맞는 방법을 사용함으로써 좋은 성능을 보일 수 있다. (그림 13)은 어떻게 임계점을 설정하는지 보여준다. BSI에서 두 기법은 비트 계산의 수에 따라 성능이 달라진다. 비트 계산은 모바일 클라이언트의 에너지 효율과 관련된다. 이 그림에서 범위 크기의 비율이 41.2%보다 작으면 A_NFRS가 A_RFNS보다 효율적이고 범위 크기의 비율이 41.2%보다 높으면 그 반대의 성능을 보인다. 따라서 우리는 질의의 제한된 지역의 범위크기에 따라서 기법을 선택할 수 있다. 본 실험에서도 범위 크기가 41.2% 이하의 질의에서는 A_NFRS를 선택하였고, 그 이외의 경우 A_RFNS를 사용하였다.

5.2 제한된 지역을 가진 최근접 탐색 기법에서의 접근 시간과 튜닝 시간

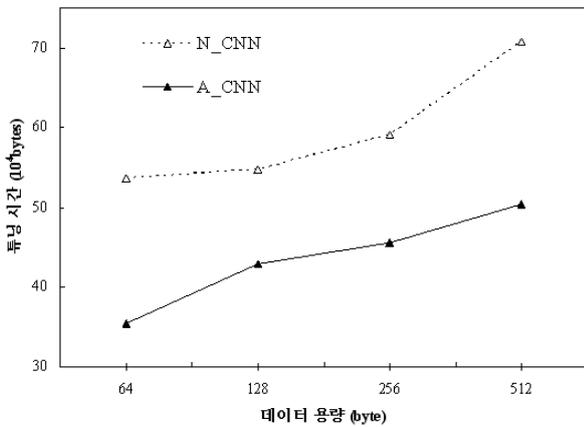
(그림 14)는 기본 기법(N_CNN)과 개선된 기법(A_CNN)의 접근 시간과 튜닝 시간을 평가한다. 각 성능은 데이터 패킷의 크기를 64byte에서 512byte까지 다양한 환경을 주어 평가한다. kNN의 k개는 10개로 설정했으며 실제 데이터를 사용했다. (그림 14(a))에서 보는 바와 같이 A_CNN은 N_CNN에 비해 접근 시간은 평균적으로 10배 이상 좋아짐을 보인다. 이는 BSI를 이용한 A_CNN이 모든 데이터를 검색하지 않고 단지 BSI만을 확인하여 모든 데이터의 위치 유무를 파악함으로써 한 번의 방송 주기 내에 질의를 처리함으로써 접근 시간을 크게 줄일 수 있다. (그림 14(b))에서는 A_CNN이 N_CNN에 비해 튜닝 시간을 평균적으로 2배 이상 좋아짐을 보인다. 이는 개선된 기법이 기본 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 튜닝 시간이 줄어 에너지 소모량을 줄인다.

5.3 다양한 제한조건을 가진 최근접 탐색 기법에서의 접근 시간과 튜닝 시간

본 절에서는 두가지 제안한 기법인 AS_CNN과 AA_CNN의 성능을 평가한다. 제안한 두 기법의 성능을 비교하기 위해 기존의 분산공간색인[9]를 비교한다. 분산공간색인(DSI)는 제약상황을 추가해 기존 [9]의 논문에서와 같이 최근접을 탐색하려 할 때 실제 데이터를 검색하여 제약상황에 맞는지 안 맞는지 검색하도록 수정하였다. 본 제한상황에서의 최근

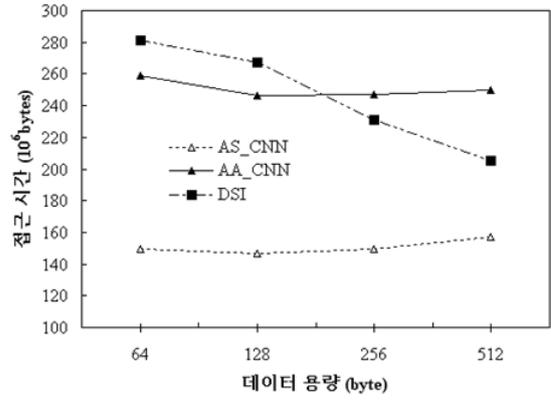


(a) 접근 시간

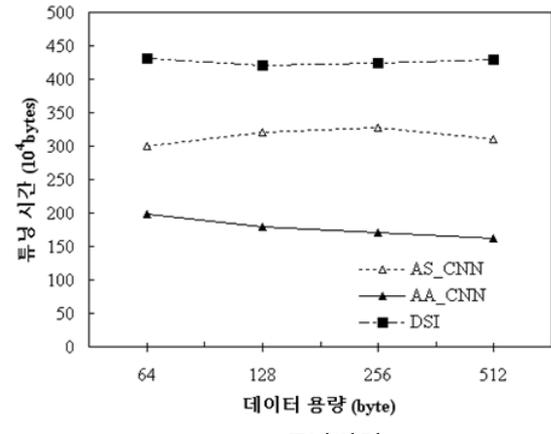


(b) 튜닝 시간

(그림 14) N_CNN과 A_CNN에서의 성능평가



(a) 접근 시간



(b) 튜닝 시간

(그림 15) AS_CNN과 AA_CNN의 성능 평가

접 질의를 위해 DSI에서는 기존의 최근접 탐색과 같은 형태로 질의를 처리한다. 하지만 최근접 질의 도중 현재 질의 점과 가장 가깝다고 계산되는 객체에 대한 실제 데이터를 검색하여 제한상황에 맞는지 여부를 검색한다. 이와 같이 진행하여 제한상황에 맞고 가장 가까운 데이터가 검색될 때까지 질의 처리를 진행한다.

(그림 15)는 두 가지 제안한 기법인 AS_CNN과 AA_CNN의 접근 시간과 튜닝 시간을 평가한다. 앞의 실험과 마찬가지로 각 성능은 데이터 패킷의 크기를 64byte에서 512byte까지 다양한 환경을 주어 평가한다. kNN의 k개는 10개로 설정했으며 실제 데이터를 사용했다. (그림 15(a))에서 보는 바와 같이 접근 시간은 AS_CNN은 AA_CNN에 비해 160% 이상, DSI에 비해 153% 이상의 좋은 성능을 보인다. 이는 AS_CNN은 후보 데이터를 순서대로 읽어오므로 하나의 방송주기 이내에 질의 처리가 끝나는 반면에 AA_CNN나 DSI는 가까운 데이터부터 읽어오므로 하나 이상의 방송주기를 필요로 할 때가 많기 때문이다. 접근 시간으로 보면 AS_CNN이 좋은 성능을 보임을 알 수 있다.

(그림 15(b))에서는 두 기법의 튜닝 시간을 측정하였다. 튜닝 시간은 AA_CNN이 AS_CNN에 비해 평균 70% 이상, DSI에 비해서는 평균 110%이상의 좋은 성능을 보임을 알

수 있다. DSI의 경우 제안한 두 기법에 비해 훨씬 많은 실제 데이터를 읽어야 되는 단점이 있어 튜닝시간은 가장 길어진다. 또한, AA_CNN의 경우 가장 가까운 데이터부터 순차적으로 검색함으로써 실제 최근접이 될 가능성이 높은 데이터를 검색하고 AS_CNN의 경우에는 최악의 경우 후보 데이터군에서 실제 최근접을 가장 나중에 찾을 수도 있어 튜닝하는 시간이 상대적으로 길어진다. 즉, AA_CNN은 실제 찾는 데이터의 수를 줄임으로서 튜닝 시간을 줄일 수 있다. 이번 실험으로서 DSI를 제외한 두 기법은 각각 접근 시간과 튜닝 시간에서 장단점을 가지고 있으며, 각 어플리케이션에 따라 상황에 맞추어 접근 시간이나 튜닝 시간에 중점을 두어 선택하여 사용할 수 있다.

6. 결론 및 향후 연구

본 논문은 무선 방송 환경에서 제한조건을 가진 경우에서의 최근접 질의(CNN)을 제안하였다. 우리는 우선 제한된 지역을 가진 경우에서 기본 기법(N_CNN)과 개선된 기법(A_CNN) 두 방법을 제안하였고 개선된 기법의 경우 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 무선기기에서 사용시 에너지와 질의의 신속성 측면에서 장점이 있다는 것을 의미한다. 또한, 다양한 제한조건을 가진 경우에 최근접 질의에 대해서 AS_CNN과 AA_CNN의 두 기법을 제안하였으며, 실험 결과 접근 시간에서는 AS_CNN이 튜닝 시간에서는 AA_CNN이 좋은 성능을 보였다.

향후 연구로는 BSI 색인 크기를 줄여 성능을 향상시킬 것이며 로드 네트워크를 적용한 도로에서의 CNN탐색으로 확장 예정이다.

참 고 문 헌

[1] Spatial datasets. 'http://www.rtreportal.org/spatial.html'.

[2] 이명수, 류병걸, 오재오, 이상근. “무선 방송 환경에서 제한된 지역의 NN 질의 처리 방법”, 한국정보처리학회 추계학술발표대회 논문집, 제15권 제2호, pp.271-274, 2008.

[3] 신한글, 이명수, 최재호, 이상근. “무선방송환경에서의 비트맵 기반 공간색인 기법”, 데이터베이스연구회지, 제23권 제1호, pp.23-36, 2007.

[4] H. Ferhatosmanoglu, I.Stanoi, D.Agrawal, and A. Abbadi. “Constrained Nearest Neighbor Queries,” In Proceedings of the International Symposium on Spatial and Temporal Databases, pp.257-276, 2001.

[5] C. Gosman and M. Lindenbaum. “On the metric properties of discrete space-filling curves,” IEEE Transactions on Image Processing, pp.794-797, 1996.

[6] A. Guttman. “R-trees: A dynamic index structure for spatial searching,” In Proceedings of the ACM SIGMOD Conference on Management of Data, pp.47-54, 1984.

[7] T. Imielinski, S. Viswanathan, and B. R. Badrinath. “Data on air organization and access,” IEEE Transactions on Knowledge and Data Engineering, Vol.9, No.3, pp.353-372, 1997.

[8] D. L. Lee, W. C. Lee, J. Xu, and B. Zheng. “Data management in location-dependent information services: Challenges and issues,” IEEE Pervasive Computing, Vol.1, No.3, pp.65-72, 2002.

[9] 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, pp.349-358, 2005.

[10] B. Zheng and D. L. Lee. “Information dissemination via wireless broadcast,” Communications of the ACM, Vol.48, No.5, pp.105-110, 2005.

[11] 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, pp.297-304, 2003.

[12] B. Zheng, W. C. Lee, and D. L. Lee. “Spatial queries in wireless broadcast systems,” Wireless Network, Vol.10, No.6, pp.723-736, 2004.



이 명 수

e-mail : lms9711@korea.ac.kr
 2004년 고려대학교 컴퓨터학과(학사)
 2006년 고려대학교 컴퓨터학과(석사)
 2006년~현 재 고려대학교 컴퓨터학과 박사
 과정
 관심분야: 센서 데이터관리, 모바일 컴퓨팅,
 위치기반 서비스 등



류 병 걸

e-mail : smart123@korea.ac.kr
 2007년 세종대학교 컴퓨터학과(학사)
 2007년~현 재 고려대학교 컴퓨터학과 석사
 과정
 관심분야: 센서 데이터관리, XML 데이터
 관리, 데이터베이스 등



오 재 오

e-mail : doublefive@korea.ac.kr
 2008년 가톨릭대학교 컴퓨터학과(학사)
 2009년~현 재 고려대학교 컴퓨터학과 석사
 과정
 관심분야: 센서 데이터관리, 모바일 컴퓨팅 등



이 상 근

e-mail : yalphy@korea.ac.kr

1994년 고려대학교 컴퓨터학과(학사)

1996년 고려대학교 컴퓨터학과(석사)

1999년 고려대학교 컴퓨터학과(박사)

2000년~2001년 University of Tokyo 특별
방문연구원

2003년~현재 고려대학교 컴퓨터학과 부교수

관심분야: Web DB, XML DB, 위치기반 정보시스템, 센서 데이
터 관리 등