

무선방송환경에서 계층적 비트맵 기반 공간 색인을 이용한 k-최근접 질의처리

송 두 희*, 박 광 진**

A Hierarchical Bitmap-based Spatial Index use k-Nearest Neighbor Query Processing on the Wireless Broadcast Environment

Doohee Song *, Kwangiin Park **

요 약

최근 무선방송 환경을 기반으로 하는 k-최근접(k-Nearest Neighbor) 질의처리가 활발히 연구되고 있다. 무선 방송환경의 장점은 서버 내에 존재하는 불특정 다수에게 일괄적으로 질의처리를 할 수 있는 확장성을 가진다는 것이다. 그러나 기존의 k-NN 질의는 무선방송환경에 적용할 경우 탐색과정에서 백트래킹이 발생하여 질의처리시간이 증가하는 단점을 가진다. 본 논문은 무선방송환경에서 k-NN 질의를 효과적으로 처리하기 위하여 계층적 비트맵 기반 공간색인(Hierarchical Bitmap-based Spatial Index: HBI)을 제안한다. HBI는 비트맵 정보와 트리 구조를 이용하여 비트맵의 크기를 줄인다. 결과적으로 방송주기를 줄임으로써 클라이언트의 청취시간과 질의처리 시간을 줄일 수 있다. 또한 비트맵 정보를 활용하여 객체의 위치를 모두 파악할 수 있기 때문에 필요한 데이터를 선택적으로 청취할 수 있다. 본 논문에서는 HBI를 k-NN 질의에 적용하여 실험을 실시하고 성능평가에서 제안 기법이 우수함을 증명한다.

▶ Keyword : k-최근접 질의, 공간색인, 위치기반 서비스, 비트맵

Abstract

Recently, k-nearest neighbors query methods based on wireless broadcasting environment are actively studied. The advantage of wireless broadcasting environment is the scalability that

• 제1저자 : 송두희 • 교신저자 : 박광진

• 투고일 : 2011. 10. 07, 심사일 : 2011. 10. 17, 게재확정일 : 2011. 10. 24.

* 원광대학교 정보통신학과(Dept. of Information and Communication, Wonkwang University)

** 원광대학교 정보통신학과(Dept. of Information and Communication, Wonkwang University), 원광대학교
공업기술개발연구소 연구위원

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0007311).

enables collective query processing for unspecified users connected to the server. However, in case existing k-NN query is applied in wireless broadcasting environment, there can be a disadvantage that backtracking may occur and consequently the query processing time is increasing. In this paper proposes a hierarchical bitmap-based spatial index in order to efficiently process the k-NN queries in wireless broadcasting environment. HBI reduces the bitmap size using such bitmap information and tree structure. As a result, reducing the broadcast cycle can reduce the client's tuning time and query processing time. In addition, since the locations of all the objects can be detected using bitmap information, it is possible to tune to necessary data selectively. For this paper, a test was conducted implementing HBI to k-NN query and the proposed technique was proved to be excellent by a performance evaluation.

▶ Keyword : k-nearest neighbor, spatial index, location based service, bitmap

I. 서론

최근 들어 무선방송환경에 관한 연구들이 활발히 진행되고 있다[1, 2, 3]. 그 이유는 무선방송환경에서 서버는 불특정 다수의 클라이언트에게 효과적으로 정보를 전송할 수 있기 때문이다. 특히 무선방송을 쉽게 이용할 수 있는 스마트폰 기기가 보급화 되면서 위치기반서비스(Location Base Service)에 대한 수요가 급격히 증가하고 있다[4, 5, 6]. 위치기반 서비스는 다양한 응용분야에서 활용된다. 그 중 대표적인 예는 k-최근접(Nearest Neighbor) 질의처리이다[7]. k-NN 질의는 질의점을 중심으로 거리가 가장 가까운 객체를 찾는 질의방법이다. 일반적으로 1개의 객체를 찾을 경우 NN 질의라고 하며, k개 이상의 객체를 찾고자 할 때 k-NN 질의라고 한다. 유선환경에서 k-NN 질의는 일반적으로 R-트리기반의 색인을 이용하여 질의를 처리한다[8]. 그러나 R-트리가 무선 환경에서 사용되면 백트래킹 문제가 발생할 수 있다[9]. 백트래킹 발생 시 질의처리 시간이 증가하게 되어 사용자의 만족도가 떨어질 뿐만 아니라 모바일 기기의 배터리 소모가 증가하게 된다. 본 논문에서는 이러한 문제점을 해결하고자 계층적 비트맵 기반 공간색인(Hierarchical Bitmap-based Spatial Index: HBI)을 제안한다. 본 논문에서 제공하는 특징과 장점을 요약하면 다음과 같다.

- HBI는 모든 객체의 좌표정보를 확인할 수 있기 때문에 효과적인 질의처리를 지원한다.
- 힐버트 커브 차수 N이 증가함에 따라 4^Nbit로 증가하는 비트맵의 단점을 보완한다.
- HBI는 트리 기법을 이용하여 전체 비트맵 중에서 비트

가 0인 부분을 대거 생략하므로 접근 시간과 청취 시간을 줄인다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 알아보고 3장에서는 HBI의 구조와 알고리즘을 설명하고, 이를 k-NN 질의에 적용시킨다. 4장에서는 실험 결과를 보여주고 5장에서 결론을 내린다.

II. 관련 연구

무선 환경에서 색인기술과 관련된 용어를 정의하면 다음과 같다.

- 대기시간 : 사용자가 질의를 보내는 시점부터 처음 색인을 수신하는데 걸리는 시간 (e)
- 접근시간 : 사용자가 질의를 보내는 시점부터 필요한 데이터를 모두 전송 받는 시점까지의 총 시간 (d)
- 청취시간 : 사용자가 요구한 데이터를 받기 위해 클라이언트가 활성화 된 상태의 총 시간 (a(색인)+b+c+e)

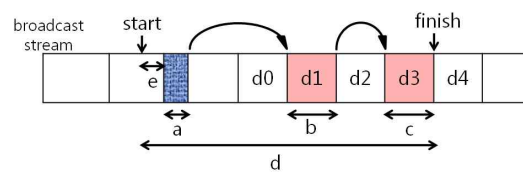


그림 1. 무선 방송 데이터의 수신
Fig. 1. Reception of wireless broadcast data

1. 비트맵

컴퓨터에서 기본적인 데이터 구조는 비트로 표기한다. 이 비트를 활용하여 비트맵으로 만드는 연구가 진행되었다[10, 11]. [11]은 시간과 공간의 트레이드오프를 개선하기 위하여

비트맵 색인을 적용시킨 기법을 제안한다.

num	B ⁸	B ⁷	B ⁶	B ⁵	B ⁴	B ³	B ²	B ¹	B ⁰
0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0	0
5	0	0	0	1	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	0

그림 2 비트맵 도표
Fig. 2. Bitmap diagram

그림 2는 일반적으로 비트맵을 표현하는 방식을 예로 보인다. num은 0부터 8까지로 가정하며, Bⁿ은 num에 따른 비트맵을 표현한다. 그러나 이 기법을 공간 질의에 적용시킬 수 있는 연구는 아직 전무한 실정이다.

2. 공간 질의 처리

기존의 색인구조는 모든 객체의 정보를 색인에 저장하는 형태를 취한다. 방송주기에서 색인이 하나일 경우 그 색인을 청취하기 위하여 접근대기시간이 증가하게 된다. 반면, 접근 시간을 줄이기 위해서 색인의 수를 증가시킬 경우 방송주기가 증가하는 단점을 가진다. 이를 보완하기 위해 분산된 색인이 제안된다[12]. 그러나 [12]의 분산된 색인 기법은 공간을 고려하지 않았기 때문에 공간 질의 처리에 적용하기 어렵다.

이후 Hilbert-Type Space-Filling Curves[13]을 이용한 공간색인 기법이 제안된다[9, 14]. [9]는 힐버트 커브를 이용한 HCI(Hilbert-Curve Index)를 제안한다. HCI는 범위 질의와 k-NN질의에 적용 가능하며 검색범위를 효과적으로 줄일 수 있는 알고리즘을 제안한다. 그러나 줄어든 범위의 시작점부터 끝점까지 모든 데이터를 청취해야 하는 단점을 가진다. [14]의 DSI(Distributed Spatial Index)는 색인마다 모든 데이터 정보를 넣지 않고 분산된 색인을 넣는다. 분산된 색인은 지수적으로(2ⁿ) 증가하고 포인터를 통하여 다음 오게 될 객체를 확인한다. 그러나 DSI는 중복된 색인을 반복적으로 검색하기 때문에 청취시간이 증가하는 단점을 가진다.

그림 3은 방송 한 주기에 데이터는 0부터 5까지 존재하며, DSI의 지수는 2²으로 증가한다고 가정한다. 방송주기에서 음영부분이 색인이며, 1번 색인이 가지고 있는 포인터 정보는 2, 3, 5번을 검색할 수 있다.

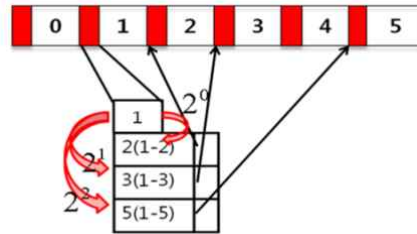


그림 3. DSI 도표
Fig. 3. DSI diagram

그림 3의 DSI는 각 데이터마다 색인을 분산시켜 삽입함으로써 접근대기시간을 감소시킨다. 그러나 색인의 지수적인 증가는 해당 색인만을 검색하기 때문에 객체 4번을 확인할 수 없다. 그래서 4번을 확인하기 위해서 추가적인 색인 청취가 발생되고 청취시간이 증가하는 단점을 가진다.

3. 비트맵 기반 공간색인

무선 방송환경에서 위치기반 서비스를 지원하는 비트맵 기반 공간색인(Bitmap-based Spatial Index)은 각 힐버트 커브 값의 데이터 유무를 비트(0, 1)로 표현한다[15]. 힐버트 커브 값에 데이터가 존재하면 비트 1, 존재하지 않을 경우 비트 0으로 전체 맵을 표현한 것을 비트맵이라고 한다.

21	22	25	26	37	38	41	42
0	0	0	0	0	0	0	0
20	23	24	27	36	39	40	43
0	0	0	0	0	0	0	0
19	18	29	28	35	34	45	44
0	0	0	0	0	0	0	0
16	17	30	31	32	33	46	47
0	0	0	0	0	0	0	0
15	12	11	10	53	52	51	48
1	0	0	0	0	0	1	0
14	13	8	9	54	55	50	49
0	0	0	0	1	0	0	0
1	2	7	6	57	56	61	62
0	0	1	0	0	0	0	1
0	1	0	0	0	0	0	0
0	3	4	5	58	59	60	63
0	1	0	0	0	0	0	0

(0001...0010₍₂₎)

그림 4. BSI 도표
Fig. 4. BSI diagram

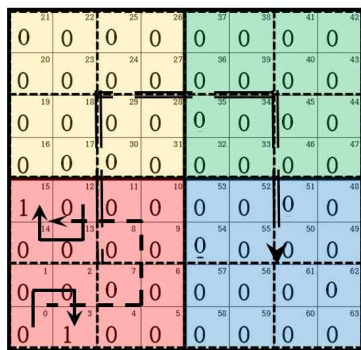
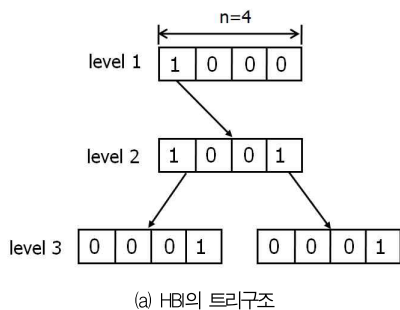
비트맵의 장점은 1비트를 통하여 1개의 객체 유무를 파악할 수 있다는 점이다. 그러나 비트맵은 힐버트 커브 차수 N이 증가함에 따라 색인의 크기가 증가하는 단점을 가진다.

III. 본 론

본 절에서는 계층적 비트맵 기반 공간색인(Hierarchical Bitmap-based Spatial Index: HBI)을 제안하고, HBI를 k-NN 질의에 적용시켜 질의 처리하는 방법을 보여준다. k-NN 질의처리에 사용되는 용어를 정리하면 다음과 같다.

- Index H : HBI 크기
- Data : 데이터 크기
- k : 찾고자 하는 객체 수
- m : 색인의 개수
- C : 대기시간
- M : k-NN 질의에서 k번째에 해당하는 힐버트 커브의 마지막 번호

1. 색인 구조



(b) 힐버트 커브 차수가 3일 때 검색순서

그림 5. 힐버트 커브 차수 3에서 HBI 검색방법
Fig. 5. HBI retrieval method in Hilbert Curve order 3

그림 5 (b)는 힐버트 커브 차수가 3인 맵이다. 그림 5 (a)

는 그림 5 (b)의 비트맵을 HBI로 변경하는 단계를 보여준다. 레벨(level) 1은 부모 노드이며, 각 레벨마다 영역 내에 객체가 존재할 경우 비트를 1로 나타내고 영역 내에 객체가 존재하지 않을 경우 비트를 0으로 나타낸다. 레벨 1은 $1000_{(2)}$ 으로 첫 번째 비트 1은 0~15번까지 한 영역을 나타내며, 두 번째 비트 0은 16~31번, 세 번째 비트 0은 32~47번, 네 번째 비트 0은 48~63번을 나타내고 있다. 레벨 2는 레벨 1에서 비트 1인 부분만을 자식 노드로 설정한다. 레벨 2는 $1001_{(2)}$ 으로 첫 번째 비트 1은 0~3번까지 한 영역으로 나타내며 두 번째 비트 0은 4~7번, 세 번째 비트 0은 8~11번, 네 번째 비트 1은 12~15번을 나타낸다. 마지막 레벨 3은 레벨 2의 첫 번째 영역과 네 번째 영역 내에 객체가 존재하는 것을 확인할 수 있다. 즉, 레벨 3에서 비트 1은 객체 3번과 15번이다. 그림 6은 기존 기법인 BSI와 새로 제안한 HBI를 그림 5(b)에 적용한 후 각각에 색인구조를 비교한다.

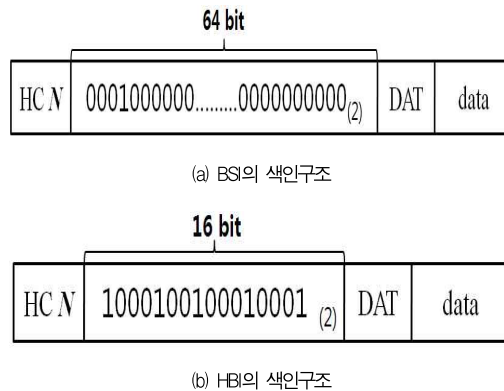


그림 6. 그림 5 (b)에 적용한 색인구조
Fig 6. Fig 5 (b) adaptive index structure

힐버트 커브 차수 3에서 BSI는 맵의 모든 객체정보를 저장하기 위하여 64bit로 표현한다. 반면에 HBI는 16bit로 맵 내에 존재하는 모든 객체를 파악할 수 있다. 그림 6은 독자이해를 돕기 위하여 간단하게 설명을 하였으나 실제로 사용되는 맵에 적용하였을 경우 힐버트 커브 차수 N이 증가하게 되어 각각에 색인 차이는 더 커진다. 또한 색인의 수(m)가 증가할 경우 BSI와 HBI의 큰 차이를 가지게 된다. HBI구조의 색인 범위는 다음과 같다.

$$4bit \leq HBI \leq \sum_{n=0}^N 4^N bit \quad (N = \text{힐버트 커브 차수 } N)$$

2. k-NN 질의

k-NN 질의는 질의점을 기준으로 거리가 가장 가까이 있는 객체 k개를 찾는 방법으로 HBI를 이용한 k-NN 질의 알고리즘과 접근시간 및 청취시간을 구하는 수식을 소개한다.

- 접근시간 = $\frac{M}{2m} \{m(Index H) + Data\} + C$
- 청취시간 = $k(Data) + Index H + C$

```

알고리즘 1. HBI를 이용한 k-NN 질의처리
Input: a query point, q, the number of nearest neighbors,
k
Output: k nearest neighbors to q
B : all objects of bit 1 in bitmap, bi: each object of bit 1,
Rbi: distance between q and bi
Procedure:
01 : compute the Hilbert Curve value of q;
02 : compute the target segments set B, each set bi
that contains object bit=1;
03 : begin initial probe and retrieve HBI;
Rbi=∞; Target=NULL; Result=NULL;
04 : while k≤KNN do
05 :   for each object Rbi do
06 :     if (arr[Rbi][1] >= 0)
07 :       temp=arr[Rbi][1];
08 :       arr[Rbi][1]=arr[Rbi][1];
09 :       arr[Rbi][1]=temp;
10 :     end if
11 :   KNN=KNN∪{Rbi};
12 : end for
13 : end while
14 : return KNN;
    
```

알고리즘 1은 HBI를 이용한 k-NN 질의처리 과정을 보여 준다. 먼저, 클라이언트는 사용자의 질의내용과 질의점을 파악한 후 HBI를 읽고 비트가 1인 객체를 후보 집합 B를 만든다. 후보 집합 B에는 후보 객체가 존재하며 후보객체와 질의점간 거리를 측정하여 거리가 짧은 순으로 정렬한다. 정렬된 후보객체 중 KNN에 해당하는 데이터를 검색한다.

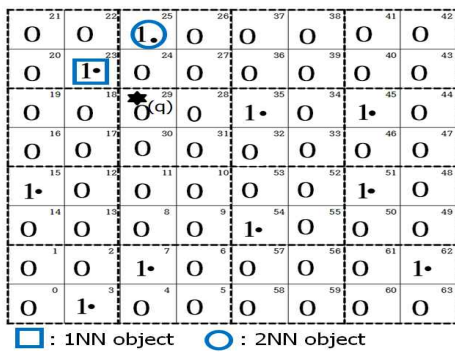


그림 7. HBI를 이용한 2NN 질의처리
Fig. 7. 2NN query processing used HBI

그림 7은 HBI를 이용한 2NN 질의처리를 보여주고 있다. 이 그림에 질의처리 과정은 먼저 질의점(28)을 파악하고 HBI를 읽은 후 맵 내에 객체가 존재하는 위치를 파악한다. 맵 내에 존재하는 객체는(3, 7, 15, 23, 25, 35, 45, 51, 54, 62)이다. 각각에 객체 위치와 질의점간에 거리를 계산하여 거리가 짧은 순으로 정렬한다(23, 25, 35, ...). 객체와 질의점간 거리가 가장 가까운 23번과 다음으로 가까운 25번이 결과 값으로 출력된다. 만약 거리가 같을 경우 객체번호가 작은 순부터 정렬된다.

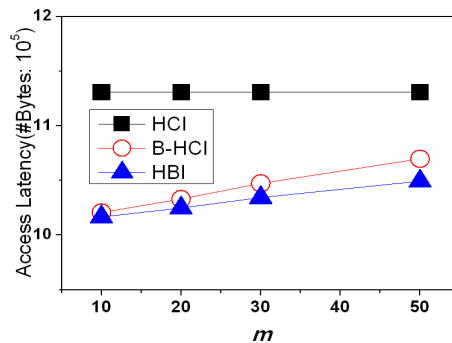
IV. 성능평가

4장에서 제안 기법인 HBI를 이용한 k-NN 질의처리에 대한 성능평가를 실시한다. 표 1은 성능평가를 위한 기본적인 파라미터 설정 값을 나타낸다.

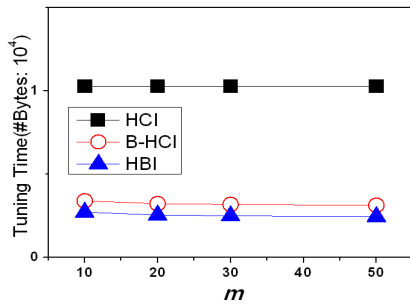
표 1. 실험 데이터 설정 값
Table 1. Experiment data set up value

항목	값
데이터 크기	1024bytes
전체비트맵	4N bits
패킷 헤더 크기	2bytes
패킷 크기	64bytes
데이터 객체수	4N개의 1/10
색인의 개수 m	10,20,30,50
k	1,3,5,10,20

본 실험에서는 색인과 데이터를 포함하여 k-NN 질의를 처리한다. 모든 실험에서 힐버트 차수는 7이며, 그림 8은 m을 변수로 하여 NN 질의에서 실험을 실시하고, 그림 9는 k를 변수로 하여 실험을 실시한다. 성능평가의 환경은 단일 서버로 가정하며, 클라이언트는 정보취득을 위해 무선 방송 채널을 사용한다.



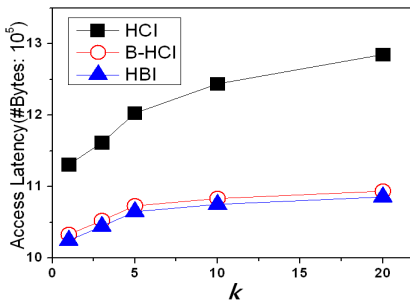
(a) 접근 시간



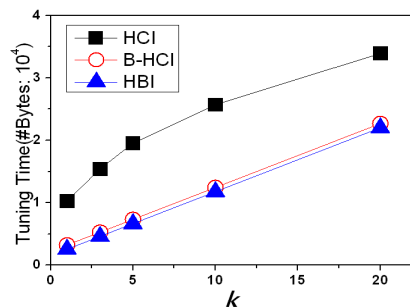
(b) 청취 시간

그림 8. NN 질의에서 m의 변화에 대한 성능평가
Fig. 8. Performance evaluation by m of change in NN query

그림 8은 NN 질의처리에서 m 을 10, 20, 30, 50개로 변화시킨 후 각각의 기법들에 대한 결과를 측정한다. 그림 8(a)는 접근시간에 대한 성능평가 결과로써 HBI가 HCI보다 평균 8.8%, B-HCI보다 평균 1.1%로 향상되었음을 확인할 수 있다. 그림 8 (b)의 청취시간 결과에서 HBI는 HCI보다 평균 75.1%, B-HCI보다 평균 21.1% 향상되었음을 확인할 수 있다.



(a) 접근시간



(b) 청취 시간

그림 9. k-NN 질의
Fig. 9. k-NN query

그림 9는 찾고자 하는 객체 k 를 1, 3, 5, 10, 20개로 변화시킨다. 그림 9(a)는 k 의 변화에 따라 각각의 기법에 대한 접근시간을 측정한다. k 가 증가함에 따라 찾고자 하는 검색범위가 증가하기 때문에 기울기가 증가한다. 그림 9(a)에서 HBI가 HCI보다 평균 12.1%, B-HCI보다 평균 0.7%로 향상됨을 보인다. 그림 9(b)의 청취시간의 경우 B-HCI와 HBI가 데이터를 선택적으로 청취하기 때문에 HCI보다 성능이 향상됨을 보이고 있다.

V. 결론

본 논문에서는 공간질의처리 시 클라이언트가 불필요한 데이터를 읽거나 필요이상의 색인 정보를 갖게 되어 접근시간과 청취시간이 증가하는 색인구조의 문제점을 보완하고자 HBI를 제안하였다.

HBI는 힐버트 커브 차수 N 이 증가함에 따라 비트맵의 색인 크기가 증가하는 단점을 줄여 접근시간을 감소시켰다. 또한 맵의 모든 객체 정보를 확인할 수 있기 때문에 사용자가 필요한 데이터만을 선택적으로 청취할 수 있게 하여 청취시간을 줄였으며, 이를 통해 클라이언트의 배터리 소모를 줄일 수 있었다. HBI는 선형적 구조를 가지기 때문에 k -NN 질의에서 백트래킹 문제의 문제점을 보완하였다.

향후 연구 방향은 복잡한 k -NN 질의를 단순화 시킬 수 있는 효과적인 질의처리 기법을 개발하고, 이를 HBI에 적용시킬 수 있는 방법을 연구할 계획이다.

참고문헌

- [1] M.-S. Chen, K.-L. Wu, and P.S. Yu, "Optimizing Index Allocation for Sequential Data Broadcasting in Wireless Mobile Computing", IEEE Trans. Knowledge and Data Eng., vol 15, no. 1, pp. 161-173, Jan./Feb. 2003.
- [2] K. Park, H. Choo, and V. Patrick, "A scalable energy-efficient continuous nearest neighbor search in wireless broadcast systems", Wireless Networks vol. 16, no. 4, pp. 1011-1031, May 2010.
- [3] D. Song and K. Park, "An Efficient Bitmap-based

- Selective Tuning Scheme for Spatial Queries in Location-based Service”, In Proc. of Intl. Conf. on Internet, pp. 503-509, Mac tan Island, Philippines, Dec. 2010.
- [4] J. Yim, K. Lee and S. Jeong, “Experimental Data Collection for Moving Pattern Information Extraction in Location based Service”, Korea Society of Computer Information, vol. 16, no. 1, pp. 17-24, June 2008.
- [5] J. Song, D. Kim and J. Lee, “Region-Based Image Retrieval System using Spatial Location Information as Weights for Relevance Feedback”, Korea Society of Computer Information, vol. 11, no. 4, pp. 1-7, Sep. 2006.
- [6] P. Bellavista, A. Kupper, and S. Helal, “Location-based Services: Back to the Future”, IEEE Pervasive Computing vol. 7, no. 2, pp. 85-89, April/June 2008.
- [7] B. Gedik, A. Singh and L. Liu, “Energy Efficient Exact kNN Search in Wireless Broadcast Environments”, 12th ACM International Workshop on Geographic Information, New York, USA, pp. 137-146, June 2004.
- [8] A. Guttman, “R-Trees: A Dynamic Index Structure for Spatial Searching”, ACM SIGMOD Record, vol. 14, no. 2, pp. 47-57, June 1984.
- [9] B. Zheng, W. C. Lee, and D. L. Lee, “Spatial Queries in Wireless Broadcast Systems”, Wireless Networks, vol. 10, no. 6, pp. 723-736, Nov. 2004.
- [10] P. O’Neil and G. Graefe. “Multi-Table Joins Through Bitmapmed Join Indices”, ACM SIGMOD Record, vol. 24, no. 3, pp. 8-11, Sep. 1995.
- [11] C. -Y. Chan, Y. E. Ioannidis, “Bitmap Index Design and Evaluation”, ACM SIGMOD Record, vol. 27, no. 2, June 1998.
- [12] T. Imielinski, S. Viswanathan, and B. R. Badri nath, “Data on Air: Organization and Access”, IEEE Transactions on Knowledge and Data Engineering, vol. 9, no. 3, pp. 353-372, May 1997.
- [13] C. Gotsman and M. Lindenbaum “On the Metric Properties of Discrete Space-Filling Curves”, IEEE Transactions on Image Processing, vol. 5, no. 5, pp. 794-797, May 1996.
- [14] B. Zheng, W. -C. Lee, Ken C. K. Lee, D. L. Lee, and M. Shao, “A Distributed spatial index for Error-Prone wireless Data broadcast”, Very Large Data Bases Journal, vol. 18, no. 4, pp. 959-986, Aug. 2009.
- [15] H. Sin, M. Lee, J. Choi and S. Lee, “Bitmap-based Spatial Index on Wireless Broadcast Environme-nts”, SIGDB-KISS, vol. 23, no. 1, pp. 23-36, April 2007.

저 자 소 개



송 두 희
 2010 : 원광대학교 정보통신학과 졸업(학사)
 현 재 : 원광대학교 대학원 정보통신학과 재학(석사)
 관심분야 : 위치기반서비스, 데이터베이스
 Email : songdohee@naver.com



박 광 진
 2000 : 고려대학교 컴퓨터학과 졸업(학사)
 2002 : 고려대학교 대학원 컴퓨터학과 졸업(석사)
 2006 : 고려대학교 대학원 컴퓨터학과 졸업(박사)
 2006~2007 : 프랑스 국립컴퓨터과학 연구소(INRIA) 박사후연구원
 현 재 : 원광대학교 전기정보통신공학부 교수
 관심분야 : 데이터베이스, 분산컴퓨팅
 Email : kjpark@wku.ac.kr

