

지오센서 네트워크의 다중 공간질의 최적화를 위한 공간질의처리비용 예측 알고리즘 연구

A Study on Cost Estimation of Spatial Query Processing for Multiple Spatial Query Optimization in GeoSensor Networks

김민수* 장인성** 이기준***
Min Soo Kim In Sung Jang Ki Joune Li

요약 최근 IoT (Internet of Things) 기술의 발전과 더불어 무선 환경에서 특정 영역에 위치하는 센서노드의 위치-센서정보를 에너지 효율적으로 수집하는 센서 네트워크 기반 공간질의처리에 대한 관심이 크게 증가하고 있다. 그리하여 센서노드에서 공간 필터링을 직접 수행하여 센서노드들 간의 통신 횟수를 감소시켜 에너지 소모를 최소화하는 다양한 공간질의처리 알고리즘 및 분산 공간색인방법들이 제안되어 왔다. 그러나 단일 공간질의처리 최적화에 중점을 두었던 기존 공간색인방법 및 알고리즘들은 IoT 환경에서 다수 사용자에게 의하여 요청되는 다중 공간질의를 최적화하여 수행하기에는 한계가 있었다. 이에 본 논문에서는 센서 네트워크에서 다중 공간질의를 에너지 효율적으로 처리할 수 있는 최적화 알고리즘을 제안하고 있다. 제안된 다중 공간질의 최적화 알고리즘은 인접 영역에 주어지는 공간질의들을 통합하여 수행하는 ‘질의통합’ 개념을 기본으로 하고 있다. 최적화 과정에서 질의들의 통합 또는 개별 수행에 대한 판단은 각 수행비용을 예측하여 결정하며, 본 논문에서는 질의처리 비용 예측 방법을 추가적으로 제안하고 있다. 끝으로, 성능평가에서는 GR-tree, SPIX, CPS의 공간색인방법에 대한 비교 실험을 통하여 제안된 알고리즘의 성능 분석결과를 제시하고 있다.

키워드 : 지오센서, 분산 공간색인, 공간질의 최적화, 공간질의처리 비용 예측

Abstract With the recent advancement of IoT (Internet of Things) technology, there has been much interest in the spatial query processing which energy-efficiently acquires sensor readings from sensor nodes inside specified geographical area of interests. Therefore, various kinds of spatial query processing algorithms and distributed spatial indexing methods have been proposed. They can minimize energy consumption of sensor nodes by reducing wireless communication among them using in-network spatial filtering technology. However, they cannot optimize multiple spatial queries which will be widely used in IoT, because most of them have focused on a single spatial query optimization. Therefore, we propose a new multiple spatial query optimization algorithm which can energy-efficiently process multiple spatial queries in a sensor network. The algorithm uses a concept of ‘query merging’ that performs the merged set after merging multiple spatial queries located at adjacent area. Here, our algorithm makes a decision on which is better between the merged and the separate execution of queries. For such the decision making, we additionally propose the cost estimation method on the spatial query execution. Finally, we analyze and clarify our algorithm’s distinguished features using the spatial indexing methods of GR-tree, SPIX, CPS.

Keywords : Geosensor, Distributed Spatial Index, Spatial Query Optimization, Cost Estimation

[†]This work was supported by the IT R&D program of MKE/KEIT. [10041790, Development of Advanced Ship Navigation Supporting System based on Oncoming International Marine Data Standard]

* Min Soo Kim, Principal Researcher, Spatial Information Research Laboratory, ETRI, minsoo@etri.re.kr (Corresponding Author)

** In Sung Jang, Senior Researcher, Spatial Information Research Laboratory, ETRI, e4dol2@etri.re.kr

*** Ki Joune Li, Professor, Dept. of Computer Engineering, Pusan National University, lik@pnu.edu

1. 서론

최근 IoT (Internet of Things) 기술의 발전과 더불어 무선 환경 기반 실시간 센서정보의 수집, 분석 및 웹 서비스 시장이 급격히 성장하고 있다. 특히 IoT의 핵심 인프라인 실시간 센서정보와 사용자 위치정보를 융합하여 서비스를 제공하는 사용자 맞춤형 LBS 시장이 크게 성장할 것으로 예측되고 있다. IoT 환경에서 이러한 센서들은 대부분 무선 환경에서 동작하며, 이로 인하여 센서정보의 실시간 수집 및 활용 과정에서 이들의 배터리 용량은 서비스의 지속성에 있어서 매우 중요한 요소가 되었다. 이에 지금까지 공간(위치)-센서정보의 수집 및 활용 과정에서 센서들의 배터리 소모량을 최소화시키기 위한 다양한 연구가 활발히 진행되어 왔다. 센서정보 수집에 있어서 에너지 효율적인 필터링을 제공하기 위한 In-network 기반 분산 색인, Aggregation 질의처리, 다중 질의 최적화 방법들이 제안되어 왔다 [12, 13]. 최근에는 특정 영역 내에 포함되는 센서노드 또는 사용자 주변에 위치한 센서노드들의 정보를 에너지 효율적으로 수집하는 지오센서 네트워크 기반의 공간질의처리 방법에 대한 관심도 크게 증가하고 있다 [5, 6, 14]. 지금까지 센서 네트워크 환경에서 동작 가능한 다양한 분산 공간질의처리 알고리즘 [7, 9, 10]과 분산 공간색인 방법들 [2, 3, 4, 8, 11]이 제안되어 왔다. 이들 방법들은 센서노드에서 직접 공간 필터링을 수행하여 센서노드 간의 무선통신 횟수를 감소시킴으로써 기존 방법에 비하여 센서노드의 에너지 소모를 감소시키는 효과를 보여 주었다.

그러나 지금까지 제안된 대부분의 지오센서 네트워크 기반 공간질의처리 방법들은 하나의 공간질의처리 과정에서의 에너지 효율성에 관한 것들이었다. 이에 본 논문에서는 다중 공간질의처리를 지오센서 네트워크에서 에너지 효율적으로 처리할 수 있는 최적화 알고리즘을 제안하고자 한다. 제안되는 다중 공간질의 최적화 알고리즘은 ‘질의통합’의 개념 [1]을 이용하는데, 질의통합은 인접한 영역에 주어진 두개 이상의 질의들을 상황에 따라서 통합하여 하나의 통합된 공간질의를 수행함으로써 센서노드의 무선통신 횟수를 감소시키는 것을 의미한다. 본 알고리즘에서 기존 질의들을 그대로 수행할 것인지 통합 질의를 수행할 것인지를 판단은 센서노드의

무선통신 횟수를 예측함으로써 결정한다. 본 논문에서는 공간질의 수행에 소요되는 무선통신 횟수를 예측하기 위한 비용 예측 모델을 제안하고 있다. 이러한 다중 공간질의 최적화 알고리즘은 IoT 환경에서 수많은 사용자들에 의하여 요청되는 위치기반 질의처리에 빈번하게 활용될 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2장 관련 연구에서는 지오센서 네트워크 기반 분산 공간색인 방법들 그리고 센서 네트워크 기반으로 제안되었던 비공간질의 최적화 방법들에 대하여 살펴볼 것이다. 3장에서는 분산 공간색인방법 기반의 다중 공간질의 최적화 알고리즘과 공간질의처리 비용 예측 방법을 제안하고, 4장에서는 다양한 분산 공간색인방법을 이용하여 제안된 다중 공간질의 최적화 알고리즘의 성능을 분석할 것이다. 끝으로 5장에서는 본 논문의 결론 및 향후 연구 방향에 대하여 살펴볼 것이다.

2. 관련 연구

지금까지 지오센서 네트워크 환경에서 효율적인 공간질의 수행을 위하여 다양한 공간색인방법들이 제안되어 왔다. R-tree MBR (Minimum Bounding Rectangle)을 이용하여 센서노드들 사이에 계층적인 분산 공간색인 트리를 구성하는 Peer-tree [2], SPIX [11], GR-tree [4] 방법들과 Quad-tree를 이용한 DQT (Distributed Quad-Tree) [3] 와 DIST (Distributed Spatio-Temporal Index) [8] 방법들이 제안되어 왔다. Peer-tree는 R-tree의 공간색인방법을 점 객체의 센서노드에 그대로 적용한 방법으로 공간색인 트리를 구성할 때, 부모와 자식에 해당되는 센서노드들이 거리에 관계없이 항상 연결이 가능하다는 가정을 하고 있다. 그러므로, 이 방법은 센서노드들 간에 대부분 근거리 멀티 홉 통신으로 연결되는 센서 네트워크 환경에서는 현실적으로 사용이 어려운 면이 있다. SPIX는 Peer-tree와 달리 센서노드들에 대하여 무선 네트워크 라우팅을 구성하고, 이러한 라우팅 정보를 기반으로 R-tree의 MBR을 이용하여 계층적인 공간색인 트리를 형성한다. 그러나 이 방법은 R-tree의 MBR 기반 클러스터링과 네트워크 라우팅을 동시에 고려해야 하기 때문에 색인 트리가 베이스스테이션에 가까워질수록 MBR들 간의 Overlap이 크게 증가하는 위험성을 가지고 있다. GR-tree는 SPIX와 유사하나 MBR

기본 분산 공간색인 트리를 구성할 때, 센서노드들 간의 무선 라우팅을 보장하면서 베이스스테이션 근처 MBR들 간의 Overlap을 최소화할 수 있는 방법을 제안하고 있다. Quad-tree를 기반으로 하는 DQT와 DIST 방법도 있으나, 이 방법들은 각 Quadrant에 리더 역할을 하는 센서노드들이 존재하고, 이들 리더 센서노드들은 다음 레벨의 리더 센서노드들과 무선 라우팅이 항상 가능함을 가정하고 있다. 그러므로 이 방법들도 Peer-tree와 마찬가지로 근거리 멀티 홉 통신으로 연결되는 센서 네트워크 환경에서는 적용이 불가능한 단점이 있다.

센서 네트워크 기반 다중 비공간질의 최적화 방법과 관련해서 Xiang et al. [13]은 다중 질의처리에서 센서노드들의 평균 무선통신 횟수를 최소화할 수 있는 TTMQO (Two-Tier Multiple Query Optimization)의 방법을 제안하였다. 이 방법은 입력으로 주어진 원본 질의들 중에서 중복된 센서정보를 요청하는 질의를 제거하여 불필요한 무선통신 횟수를 감소시키는 ‘질의 재작성’ 방법을 이용한다. Trigoni et al. [12]은 다중 집계 질의 최적화 방법을 제안하였는데, 이 방법은 집계 질의 처리에 필요한 센서정보를 다수 질의가 상호 공유하게 함으로써 무선통신 횟수를 감소시키도록 하고 있다.

지금까지 지오센서 네트워크 환경을 위한 다양한 분산 공간색인 방법과 다중 질의처리 최적화 방법들이 제안되어 왔으나, 다중 공간질의 최적화에 대한 연구가 부족하였다. 이에 본 논문에서는 기존 분산 공간색인 방법을 이용하여 다중 공간질의 최적화를 수행할 수 있는 알고리즘을 제안하고자 한다.

3. 분산 공간색인 기반 다중 공간질의 최적화 알고리즘

3.1 다중 공간질의 최적화 알고리즘

본 논문에서 제안되는 다중 공간질의 최적화 알고리즘은 기본적으로 인접한 영역에 주어지는 두개 이상의 유사한 질의들을 통합하여 더 적은 수의 새로운 질의 집합을 생성하여, 이를 수행하는 ‘질의통합’ 수행의 개념을 이용하고 있다. 다시 말하면, 가능한 더 적은 수의 통합 공간질의들을 수행함으로써 센서노드에서 발생하는 무선통신 횟수를 감소시키는 것을 의미한다. 이러한 질의통합 수행 기반의 최적화 알고리즘은 당연히 질의들의 통합 수행 및

개별 수행 중에서 어떤 방법이 더 이익이 되는지 판단할 수 있어야 한다.

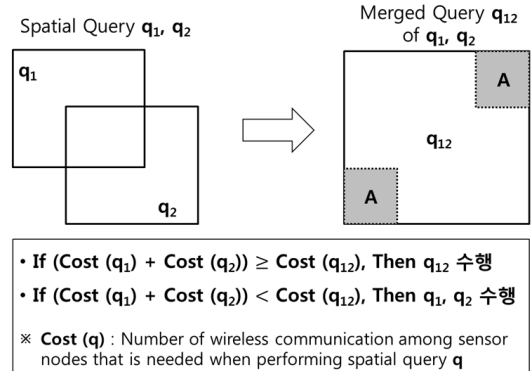


Figure 1. Merged execution of spatial queries

예를 들면, 그림 1에서 보는 바와 같이 최적화 알고리즘은 공간질의 q_1, q_2 를 개별적으로 수행할지 아니면 통합질의 q_{12} 를 수행할지를 판단하기 위한 $Cost(q_i)$ 를 예측할 수 있어야 한다. 일반적으로 q_1, q_2 를 개별적으로 센서노드에 보내어 질의를 수행하는 것보다 통합질의 q_{12} 를 수행하는 것이 센서노드들의 무선통신 횟수를 훨씬 더 감소시킬 것으로 생각된다. 그러나 q_{12} 는 그림 1의 영역 A와 같이 불필요한 공간질의 영역을 추가로 포함하게 되고, 이로 인하여 무선통신 횟수가 오히려 증가하게 되는 위험도 가지고 있다. 이에 본 논문에서는 공간질의 q_i 의 수행에 소요되는 $Cost(q_i)$ 를 예측하기 위한 비용 모델을 제안하며, 이러한 비용 모델 기반의 다중 공간질의 최적화 알고리즘은 다음과 같다.

본 논문의 최적화 (Optimization) 알고리즘은 기존에 반복적으로 수행되고 있는 공간질의 리스트 S에 대하여 신규 질의 q_{in} 가 입력되었을 때의 q_{in} 을 최적화하여 수행하는 과정을 보여준다. 만약 q_{in} 이 입력되었을 때 S가 NULL인 경우, q_{in} 은 S에 바로 삽입되고 수행된다 (라인 1-2). 그렇지 않은 경우, S에 속한 개별 질의들인 q_s 에 대하여 q_{in} 의 통합질의 수행비용과 q_s 와 q_{in} 의 개별 질의 수행 비용을 비교하여 최소 비용이 소요하는 경우를 찾는다 (라인 6-16). 최소 비용의 통합 공간질의 $optQuery$ 가 발생한 경우, 기존의 q_s 를 S에서 삭제하고 $optQuery$ 를 S에 새로이 삽입하여 실제 질의 리스트 S를 수행한다 (라인 17-19).

Table 1. Variables

Algorithm: Optimization (q_m, S)
(입력 공간질의 q_m , 통합 공간질의 리스트 S)

```

1 If ( $S == \text{NULL}$ ) Then
2    $S.\text{insert}(q_m)$ ;
3 Else
4    $i \leftarrow 1$ ;  $\text{cost1} \leftarrow \text{Cost}(q_m)$ ;
5    $\text{optQuery} \leftarrow \text{NULL}$ ;
6   For a query  $q_s$  of  $S[i]$  Do
7      $q_m \leftarrow \text{Merge}(q_m, q_s)$ ;  $m\text{Cost} \leftarrow \text{Cost}(q_m)$ ;
8      $\text{cost2} \leftarrow \text{Cost}(q_s)$ ;
9     If ( $(\text{cost1} + \text{cost2}) \geq m\text{Cost}$ ) Then
10      If ( $\text{optCost} > m\text{Cost}$ ) Then
11         $\text{optIndex} \leftarrow i$ ;  $\text{optCost} \leftarrow m\text{Cost}$ ;
12         $\text{optQuery} \leftarrow q_m$ ;
13      End If
14    End If
15     $i \leftarrow i + 1$ ;
16  End For
17 If ( $\text{optQuery} \neq \text{NULL}$ ) Then
18    $S.\text{delete}(\text{optIndex})$ ;  $S.\text{insert}(\text{optQuery})$ ;
19 End If
20 End If

```

Variables	Descriptions
q	A spatial query which is sent to nodes
$\text{Cost}(q)$	Total wireless communication cost
$\text{Cost}_P(q)$	Wireless communication cost of a query propagation
$\text{Cost}_A(q)$	Wireless communication cost of sensor result acquisition
N	Total number of sensor nodes
ht	Height up to QA in a spatial index tree
d	Distance between BS and QA
d_i	Distance between BS and ic
$\text{Area}(QA)$	Area of QA
$\text{Area}(W)$	Total area where sensors are deployed
sn	Expected number of sensors within QA
QB	Area where intermediate sensor may be located, when sending sensor data within QA
ht'	Height up to QB in a spatial index tree
e	Distance between BS and QB
e_i	Distance between BS and ic

3.2 분산 공간색인 기반 공간질의 비용 예측 방법

본 논문에서 제안된 다중 공간질의 최적화 알고리즘은 센서노드들의 무선통신 비용인 $\text{Cost}(q)$ 의 예측을 필요로 한다. 그러나 임의의 공간질의 q 를 센서 네트워크로 전송하지 않고 베이스스테이션 (BS)이 보유한 정보만을 이용하여 이러한 $\text{Cost}(q)$ 를 예측하는 것은 쉽지 않다. 이에 본 연구에서는 트리 형식의 분산 공간색인 방법을 이용한 $\text{Cost}(q)$ 예측 방법을 제안하고 있다. 지오센서 네트워크에서 공간질의 q 의 수행을 위해서는 기본적으로 q 의 센서노드로 전송비용 ($\text{Cost}_P(q)$)과 센서정보 수집비용 ($\text{Cost}_A(q)$)이 필요하기 때문에, $\text{Cost}(q)$ 는 다음 수식 (1)과 같이 예측된다.

$$\text{Cost}(q) = \text{Cost}_P(q) + \text{Cost}_A(q), \quad (1)$$

본 연구의 $\text{Cost}(q)$ 예측에서 BS는 지오센서 네트워크에 분산 공간색인 트리가 어떻게 구성되어 있는지 전혀 알 수 없기 때문에, 일반적으로 알 수 있는 전체 센서노드 수 (N)와 센서노드의 무선통신 거리 (c) 정보만을 알고 있다고 가정하였다.

본 논문에서 이용되는 수식 내에 사용되는 주요 변수들을 요약하면 표 1과 같다.

3.2.1 공간질의 전송비용 ($\text{Cost}_P(q)$) 예측

일반적으로 공간질의는 BS에서 분산 공간색인 트리의 라우팅을 통하여 질의영역에 포함되는 센서노드들에 전송된다. 그러므로, BS로부터 해당 센서노드들까지에 해당되는 공간색인 트리 높이만큼의 무선통신 횟수를 기본적으로 필요로 하게 된다. 또한, 하나의 센서노드가 아니라 질의영역에 포함되는 다수의 센서노드들에게 공간질의가 전송되어야 하기 때문에 색인 트리를 따라 다중 라우팅 경로가 필요하기도 하다. 본 논문에서는 BS와 질의영역 사이에 이러한 다중 라우팅 경로가 얼마나 많이 필요할지를 Heuristic하게 예측함으로써, $\text{Cost}_P(q)$ 를 예측하고 있다. 구체적으로 분산 공간색인 트리에서 Root 역할을 하는 BS에서 질의 q 가 처음 전파될 때는 하나의 라우팅 경로면 충분하다. 그러나 q 가 전체 질의영역 (QA)에 가까워질수록 점점 더 많은 수의 다중 라우팅 경로가 필요하게 된다.

그림 2에서 보듯이, $\text{Cost}_P(q)$ 예측은 BS와 QA 사이에 임의의 공간색인 트리 (T)가 있다고 가정할 때, 색인 트리의 폭이 센서노드들의 기본 통신 거리의 배수가 되는 지점들, $c, 2c, \dots, nc$ 를 찾는다. 이후, 각 지점별로 분리하여 필요한 최소한의 라우팅 경로의 수를 예측하는데, 본 논문에서는 BS부터 시작하여 차례대로 최소 1, 2, ..., n 개의 라우팅 경로가 필요할 것이라고 예측한다. 다시 말하면 최소

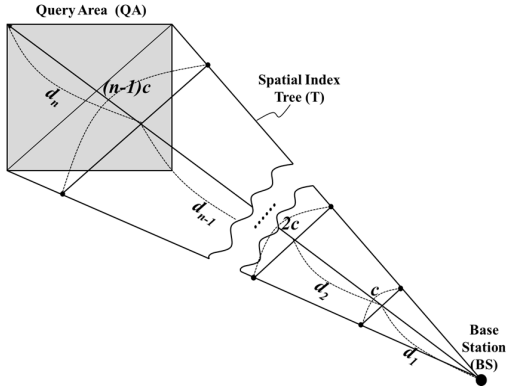


Figure 2. Cost estimation of query propagation

1, 2, ..., n개의 라우팅 경로를 통하면 질의 q가 최종적으로 QA 내에 존재하는 모든 센서노드들에게 전송될 수 있다고 판단하는 것이다. 그러므로, 전체 색인 트리의 높이를 ht라고 하고, c, 2c, ..., nc의 각 지점까지의 거리를 d1, d2, ..., dn라고 하면, CostP(q)는 다음 수식 (2)와 같이 예측된다.

$$Cost_P(q) = \sum_{i=1}^n \left((ht \times \frac{d_i}{d}) \times i \right), \quad (2)$$

여기서 d는 BS에서 QA까지의 최대거리이며, i는 라우팅 경로의 수를 의미하며, ht를 d/c로 간단히 예측한다면 CostP(q)는 다음과 수식 (3)과 같이 계산된다.

$$Cost_P(q) \cong \sum_{i=1}^n \left(\frac{d_i}{c} \times i \right), \quad (3)$$

3.2.2 센서정보 수집비용 (CostA(q)) 예측

센서정보 수집비용을 정확히 예측하기 위해서는 우선 QA 내에 속한 센서노드들의 수 (sn)를 파악해야 하며, 이후 그들이 보유한 센서정보의 전송비용을 파악해야 한다. 본 논문에서는 센서노드들이 전체 영역 (W)에 랜덤하게 분포되어 있다고 가정하고 있으며, sn 예측은 간단하게 W의 면적 대비 QA의 면적 비율을 이용하고 있다.

$$sn = N \times \frac{Area(QA)}{Area(W)}, \quad (4)$$

센서정보 수집비용 (CostA(q))은 공간색인 트리의 다중 라우팅 경로를 이용한 공간질의 전송비용과 유사한 방식으로 예측할 수 있다. 단지, 센서정보 수집 시에는 QA에 속한 모든 센서노드들이 자신의 정보를 전송하기 위하여 초기 1회의 무선통신 비용이 필요하다고 예측한다. 1회의 무선통신 이후에 센서정보들은 공간색인 트리 상에 위치할 것으로 예측되는데, 특히 그림 3에서 보는 바와 같이 QA에 속하지만 가능하면 BS에 가까운 QB 영역 부근에 위치할 것으로 예측한다. 끝으로 QB 영역으로부터 BS까지의 센서정보 수집비용은 공간질의 전송비용 CostP(q)와 동일한 방식으로 예측이 가능하며, 다음 수식 (5)와 같이 나타낼 수 있다.

$$Cost_A(q) = sn + \sum_{i=1}^n \left((ht' \times \frac{e_i}{e}) \times i \right), \quad (5)$$

$$\cong sn + \sum_{i=1}^n \left(\frac{e_i}{c} \times i \right)$$

여기서, e는 BS에서 QB 영역까지의 최단거리를 의미하며, i는 라우팅 경로의 수를 의미하며, ht'은 BS에서 QB 영역까지의 공간색인 트리의 높이로 e/c로 간단히 예측한다.

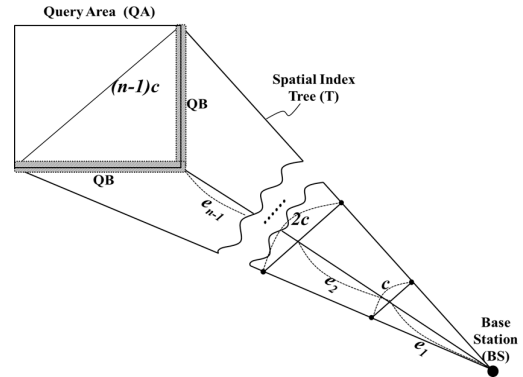


Figure 3. Cost estimation of sensors acquisition

4. 성능 평가 및 분석

4.1 실험 환경

본 논문에서 제안된 분산 공간색인 방법 기반 다중 공간질의 최적화 알고리즘에 대한 성능 평가 및

분석에서는 지오센서 네트워크에서 동작 가능한 대표적인 GR-tree, SPIX, 그리고 CPS (Closest Parent Selection)의 공간색인 방법을 이용하여 성능을 비교 분석하였다. 여기서 CPS 방법은 분산 공간색인 트리 생성과정에서 자식 센서가 부모 센서를 선택할 때, 연결 가능한 센서 중에서 베이스스테이션에 가장 가까운 센서를 일차적으로 부모로 선택하는 단순한 색인 구성 방법으로 쉽게 구현하여 활용이 가능하다. 본 실험의 성능 비교 기준으로는 GR-tree, SPIX, CPS 방법에 대하여 제안된 다중 공간질의 최적화 알고리즘을 적용하였을 때 발생하는 $Cost(q)$ 에서의 이득을 이용하였다. 본 실험은 시뮬레이션을 통하여 수행되었으며, 시뮬레이션에 이용된 실험 환경은 표 2와 같다. 끝으로, 본 실험에 이용된 다중 공간질의 최적화 알고리즘, 질의처리 비용 예측 방법과 GR-tree, SPIX, CPS의 분산 공간색인 방법은 Windows 환경에서 JDK 1.6을 이용하여 구현되었다.

Table 2. Experimental setup

Environments	Descriptions
Total area (W)	• Wors space: 2,000m x 2,000m
Sensor node	• No. of nodes(N): 2,000~20,000 • Communication distance(c): 100m • Packet size(p): 128 bytes
Spatial Query (q)	• Type of spatial query: MBR • No. of spatial query: 50~500 • Area of spatial query: 50m~200m • Distribution of spatial query - Random distribution - Skewed distribution to 1/4 area
Spatial Index	• GR-tree, SPIX, CPS

4.2 실험 결과 및 분석

본 성능평가 실험에서는 먼저 다중 공간질의 처리 최적화를 위하여 제안된 $Cost(q)$ 예측방법이 GR-tree, SPIX, CPS의 공간색인 방법에 효율적으로 적용 가능한지를 검증하였으며, 이후, 각 공간색인 방법에 대하여 제안된 다중 공간질의처리 최적화 알고리즘의 성능을 분석하였다.

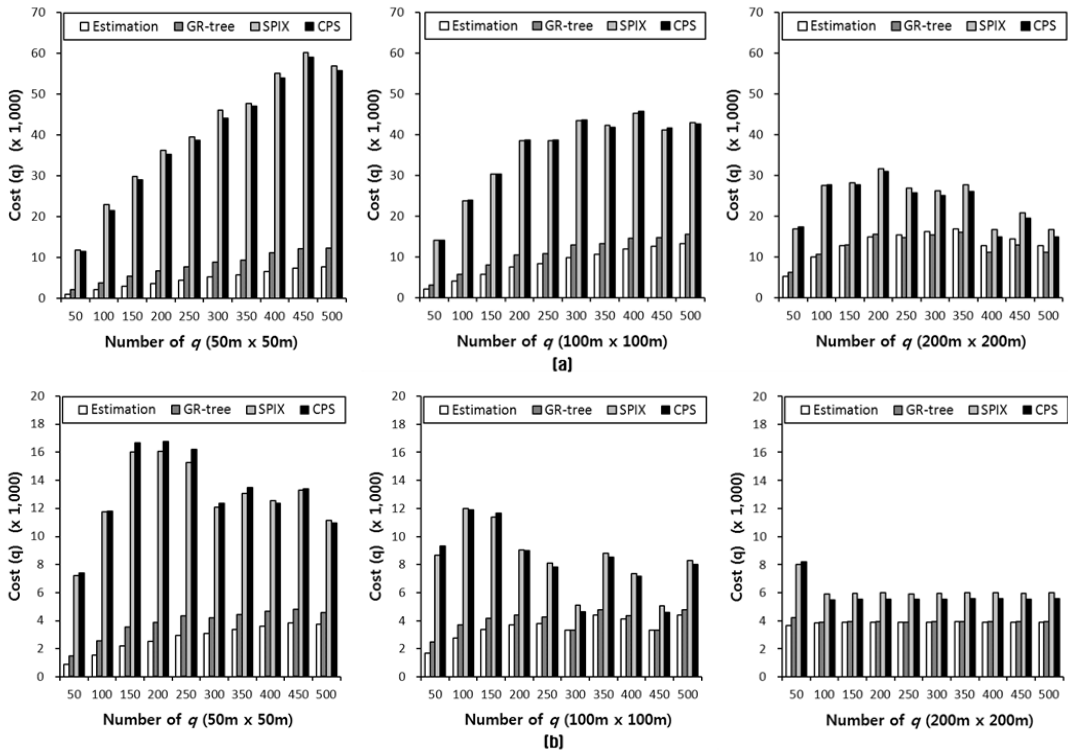


Figure 4. Comparison of real execution cost and estimation cost on GR-tree, SPIX, CPS with varying number and size of queries: (a) Random distribution and (b) Skewed distribution of queries

4.2.1 다중 공간질의 처리비용 예측 방법 검증

본 실험에서는 GR-tree, SPIX, CPS 방법에 대하여 다양한 실험을 통하여 $Cost(q)$ 에 대한 예측값과 실제값을 비교 분석하였는데, 실험 1에서는 공간질의 q 의 수를 50개~500개로, q 의 영역을 50m, 100m, 200m로 다양하게 변동하면서 질의를 수행하였을 때 예측된 $Cost(q)$ 와 각 공간색인방법 기반 q 의 실제 수행 $Cost(q)$ 를 비교 분석하였다. 여기서, 전체 영역 W 에 기본적으로 설치된 센서노드 수 N 은 10,000, 센서노드들의 무선통신 거리 c 는 100m, 패킷 크기 p 는 128 bytes로 설정하였다. 그림 4는 실험 1에 대한 결과로서, 구체적으로 그림 4(a)와 4(b)는 각각 q 가 Random 또는 Skewed 분포되었을 때의 결과를 보여주고 있다. 그림 4에서 Estimation을 통한 $Cost(q)$ 와 GR-tree 기반 실제 질의 수행 $Cost(q)$ 를 보면 q 의 수, 영역, 그리고 분포에 관계없이 유사한 패턴 및 값을 보여줌을 알 수 있다. 반면에 SPIX와 CPS 기반의 $Cost(q)$ 는 Estimation 비용과 큰 차이를 보이고 있으며, q 의 영역이 200m x 200m으로 크게 증가한 경우에만 그 차이가 다소 감소함을 볼 수 있다.

실험 2에서는 실험 1과 달리 q 의 영역을 100m로 고정하고 대신 센서노드의 수 N 을 2,000개~20,000개로, q 의 수를 100개, 300개, 500개로 다양하게 변동하면서 $Cost(q)$ 를 비교 분석하였으며, 실험 결과는 그림 5와 같다. 그림 5에서도 그림 4와 마찬가지로 GR-tree 기반 $Cost(q)$ 는 N 과 q 의 수에 관계없이 Estimation 비용과 유사한 값 및 패턴을 보여주고 있다. 이에 비해, SPIX와 CPS 방법은 N 과 q 의 수에 비례하여 $Cost(q)$ 에서 큰 차이를 보여주고 있다. 다만, 300개 이상의 많은 질의가 Skewed하게 분포되어 공간질의 최적화 알고리즘이 효과적으로 적용되어 $Cost(q)$ 가 크게 감소하는 경우에만 그 차이가 감소함을 볼 수 있다.

실험 1과 2의 결과를 종합해보면, 제안된 공간질의 처리비용 예측 방법이 공간색인 트리를 기반으로 만들어 졌다고 하더라도 GR-tree, SPIX, CPS와 같이 색인 트리의 생성방법에 따라 $Cost(q)$ 에서 큰 차이가 발생함을 알 수 있다. 구체적으로 SPIX는 MBR 클러스터링 기반 센서노드들 간의 공간 인접성을 유지하는 방식으로 색인 트리를 생성하며, CPS는 인접성 보다는 베이스스테이션에 가장 가까

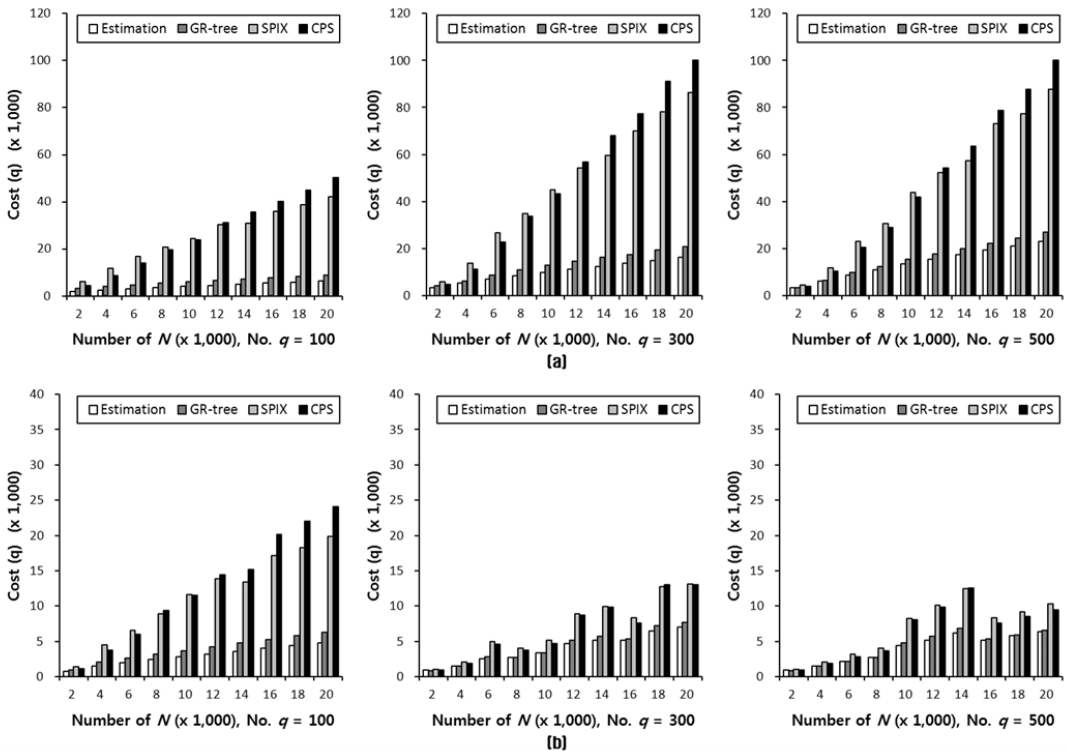


Figure 5. Comparison of real execution cost and estimation cost on GR-tree, SPIX, CPS with varying number of sensor nodes: (a) Random distribution and (b) Skewed distribution of queries

운 노드를 부모로 설정하는 최단거리 라우팅 기반의 색인 트리를 생성하는데, 비해 GR-tree는 Grid 기반의 클러스터링을 통한 공간적인 인접성 보장과 동시에 동일 Grid 내에서의 최단거리 라우팅 기반의 색인 트리를 생성하는 특징을 가지고 있다. 본 논문에서 제안된 비용 예측방법은 기본적으로 최단거리 라우팅의 특징을 기본적으로 이용하고 있으나 공간 인접성을 고려하여 다중 라우팅 경로의 수를 예측하기 때문에 결론적으로 위의 실험 1과 2에서 GR-tree가 제안된 예측방법과 가장 유사한 $Cost(q)$ 를 보여주는 것이 타당하다고 판단된다.

4.2.2 다중 공간질의처리 최적화 방법 성능 분석

본 실험에서는 GR-tree, SPIX, CPS의 공간색인 방법에 대하여 제안된 다중 공간질의처리 최적화 알고리즘의 적용 전후의 $Cost(q)$ 를 비교 분석하였다. 구체적으로 각 공간색인방법에 대하여 다음 네 경우에 대한 질의처리 비용을 조사함으로써 제안된 최적화 방법의 성능을 비교 분석하였다.

- All Query Processing (All-QP): 공간질의 q 를 최적화하지 않고 개별적으로 모두(All) 수행하였

을 때의 질의처리 비용

- Estimation (EST): 공간질의 q 를 최적화시킨 이후, 제안된 공간질의 처리비용 예측 방법에 의하여 예측된 질의처리 비용
- Optimized QP (Opt-QP): 제안된 공간질의 최적화 알고리즘을 이용하여 공간질의 q 를 수행하였을 때의 질의처리 비용
- One-QP: 공간질의 q 를 개별적으로 또는 최적화하여 수행하지 않고, 모든 센서노드들의 정보를 수집하기 위하여 전체 영역 W 를 이용하여 단 하나의 질의만을 수행하였을 때의 질의처리 비용

실험 3에서는 10,000 개의 센서노드에 100m 크기의 공간질의 q 를 50개~500개로 다양하게 변동시키면서 GR-tree, SPIX, CPS 방법에 대하여 질의를 수행하였을 때 제안된 다중 공간질의 최적화 방법의 성능을 분석하였다. 먼저 공간질의를 모두 수행해야 하는 All-QP 방법은 GR-tree, SPIX, CPS의 공간색인 종류에 관계없이 q 의 수에 비례하여 $Cost(q)$ 가 증가함을 짐작할 수 있는데, 그림 6에서 공간색인의 종류와 공간질의의 분포에 따라 증가폭에서는 다소 차이가 있지만 All-QP 방법의 $Cost(q)$ 가

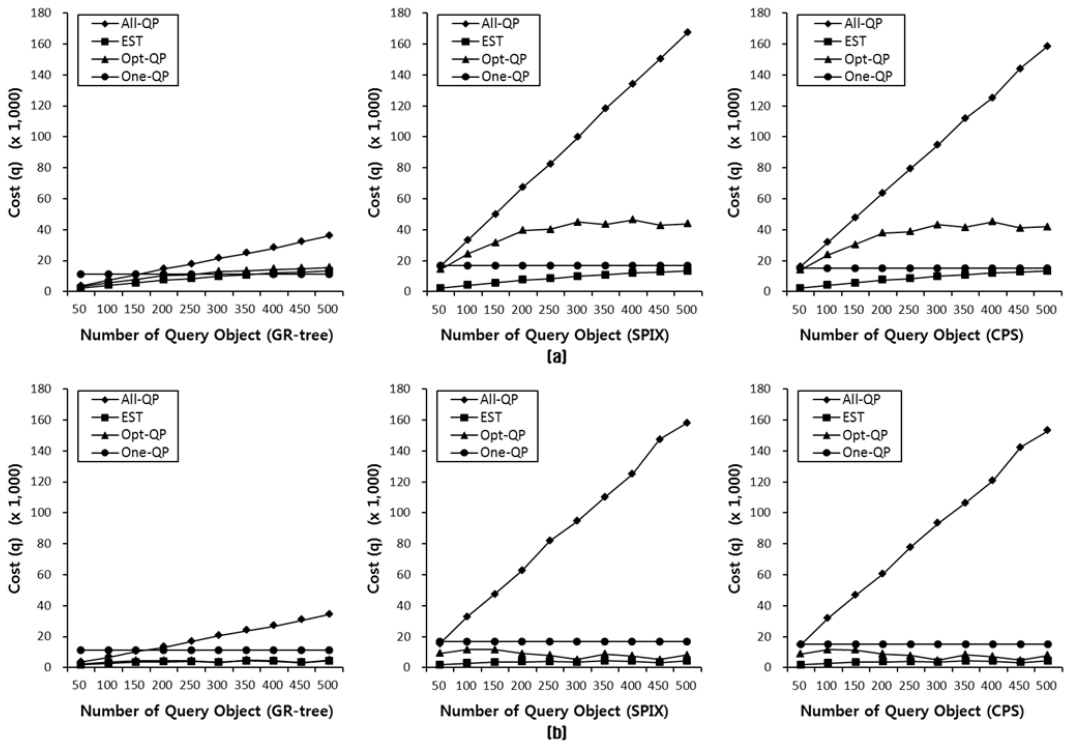


Figure 6. Performance comparison of multiple spatial query optimization on GR-tree, SPIX, CPS with varying number of queries: (a) Random distribution and (b) Skewed distribution of queries

비례적으로 증가하고 있음을 볼 수 있다. 이에 비하여 EST 방법과 Opt-QP 방법을 보면 All-QP 방법 대비, $Cost(q)$ 에 있어서 효율적일뿐만 아니라 q 의 개수 증가에도 불구하고 $Cost(q)$ 에 큰 변화가 없음을 볼 수 있다. 이는 제안된 최적화 알고리즘이 상황에 따라 차이는 있겠지만 적용이 가능하다는 것을 의미한다. 그림 6에서 추가로 주목할 점은 많은 수의 q 가 Random하게 분포될 때는 One-QP 방법이 오히려 제안된 Opt-QP 방법보다 효율적일 수도 있다는 것이다. 이는 q 의 수가 크게 증가하여 질의 영역이 전체 영역 W 에 근접하는 경우에는 다수 공간질의를 최적화하여 수행하는 것 보다는 W 에 대하여 단 하나의 공간질의를 수행하는 것이 더욱 효과적일 수 있기 때문이다. 실제로 q 가 Random하게 분포된 그림 6(a)에서 SPIX, CPS 공간색인의 경우 대부분 One-QP 방법이 Opt-QP 방법보다 효율적인 성능을 보여주고 있다.

실험 4에서는 실험 3과 다르게 q 의 수를 중간인 250개로 고정하고, 대신 센서노드의 수 N 을 2,000개~20,000개로 변동시키면서 질의를 수행하였을 때

제안된 다중 공간질의 최적화 방법의 성능을 분석하였으며, 실험결과는 그림 7과 같다. 그림 7을 보면 공간색인의 종류와 공간질의의 분포에 따라 $Cost(q)$ 증가폭에서 다소 차이는 있으나, N 의 증가에 따라 All-QP 방법이 EST, Opt-QP, One-QP 방법에 비하여 높은 $Cost(q)$ 를 소모함을 볼 수 있다. 또한, 실험 3에서와 마찬가지로 많은 수의 q 가 Random하게 분포된 그림 7(a)를 보면 SPIX와 CPS 색인은 One-QP 방법이 Opt-QP 방법보다 $Cost(q)$ 에서 효율적임을 볼 수 있다. 반면에 그림 7(b)와 같이 q 가 Skewed하게 분포될 때는 공간색인의 종류 및 N 의 증가와 관계없이 Opt-QP 방법이 One-QP 방법보다 더 적은 $Cost(q)$ 를 소모함을 볼 수 있다.

실험 3과 4의 결과를 종합해보면 다양한 공간색인방법에 대하여 제안된 Opt-QP 방법이 All-QP 방법보다 $Cost(q)$ 에서 항상 우수한 성능을 보여줄 수 있었다. 그러나 SPIX와 CPS 색인의 경우는 다수의 공간질의가 W 에 넓게 분포되는 경우, Opt-QP 방법이 All-QP 방법보다는 성능이 우수하

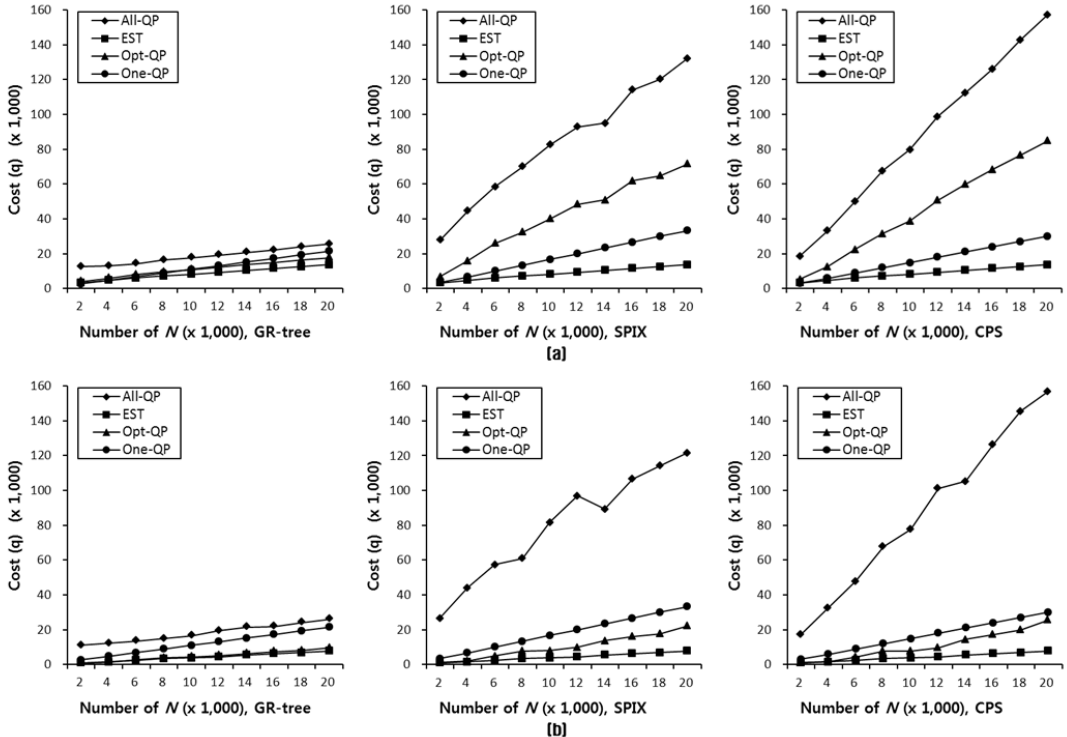


Figure 7. Performance comparison of multiple spatial query optimization on GR-tree, SPIX, CPS with varying number of sensor nodes: (a) Random distribution and (b) Skewed distribution of queries

나 One-QP 방법보다 성능이 떨어지는 경우를 볼 수 있었다. 그러므로, 제안된 공간질의 최적화 알고리즘은 GR-tree 색인에 대해서는 바로 적용이 가능하다고 판단되나, SPIX와 CPS 색인에 대해서는 많은 수의 공간질의가 광범위하게 분포되는 경우에는 질의 수 및 질의 영역을 신중하게 고려하여 적용할 필요가 있는 것으로 보인다.

5. 결론

본 논문에서는 지오센서 네트워크 환경에서 다중 공간질의를 에너지 효율적으로 처리할 수 있는 최적화 알고리즘을 제안하였다. 제안된 최적화 알고리즘은 인접 영역에 주어지는 다수의 공간질의들을 통합하여 수행하는 질의통합 개념을 이용하고 있다. 또한, 최적화 과정에서 공간질의들의 통합기준 판단을 위하여 분산 공간색인 트리를 이용한 공간질의 처리비용 예측 방법을 제안하였다. 특히, 이러한 질의처리 비용 예측은 베이스스테이션에서 수행되며, 전체 센서노드의 수와 센서의 무선통신 거리의 한정된 정보만을 이용하여 예측이 가능한 특징을 가지고 있다. 성능평가 및 분석에서는 GR-tree, SPIX, CPS 등의 다양한 공간색인방법에 대하여 다중 공간질의처리 비용 예측 방법을 검증하고, 각 공간색인방법별로 제안된 다중 공간질의처리 최적화 알고리즘의 성능을 분석하였다. 실험결과를 요약하면 GR-tree 색인은 정확한 공간질의 처리비용 예측이 가능하고 최적화 알고리즘의 성능도 매우 우수하였다. 하지만, SPIX, CPS 색인은 다소 부정확한 질의 처리비용 예측과 다수의 공간질의가 광범위하게 분포하는 경우, 최적화 알고리즘의 성능이 떨어지는 위험성이 있었다. 이는 제안된 공간질의 처리비용 예측 방법이 GR-tree, SPIX, CPS와 같이 색인 트리의 생성방법에 따라 $Cost(q)$ 에서 큰 차이가 발생하였기 때문이다. 구체적으로 SPIX는 MBR 클러스터링 기반으로 센서노드들 간의 공간 인접성을 유지하는 방식으로 색인 트리를 생성하며, CPS는 베이스스테이션에 가장 가까운 노드를 부모로 설정하는 최단거리 라우팅 기반으로 색인 트리를 생성하며, GR-tree는 Grid 기반 클러스터링을 통한 공간적인 인접성과 동일 Grid 내에서의 최단거리 라우팅을 고려하여 색인 트리를 생성하는데 비하여, 본 논문의 제안된 비용 예측방법은 기본적으로 최단거

리 라우팅의 특징을 이용하며 다중 라우팅 경로의 수 예측에서는 공간 인접성을 이용하였기 때문이다. 이에 향후 연구로는 본 논문에서 제안된 방법을 보완하여 SPIX, CPS 등의 다양한 공간색인 방법에 대해서도 효율적으로 적용 가능한 최적화된 질의 처리비용 예측 방법이 필요할 것으로 생각된다.

끝으로, 이러한 다중 공간질의의 최적화된 수행방법은 미래 IoT 환경에서 스마트폰을 사용하는 수많은 사용자들에 의하여 요청되는 위치기반 질의처리에 효율적으로 활용될 수 있을 것으로 기대된다.

참고문헌

- [1] Crespo. A. 2003, Query merging: improving query subscription processing in a multicast environment, IEEE Transaction on Knowledge and Data Engineering, 15(1):174-191.
- [2] Demirbas. M; Ferhatosmanoglu. H. 2003, Peer-to-Peer Spatial Queries in Sensor Networks, Proc. 3rd Int'l Conference on Peer-to-Peer Computing, 32-39.
- [3] Demirbas. M; Lu, X. 2007, Distributed Quad-Tree for Spatial Querying in Wireless Sensor Networks, Proc. IEEE Int'l Conference on Communications, 3325-3332.
- [4] Kim. M. S; Jang. I. S. 2011, The GR-tree: An Energy-Efficient Distributed Spatial Indexing Scheme in Wireless Sensor Networks, Journal of Korea Spatial Information Society, 19(5):63-74.
- [5] Kim. M. S; Kim. J. W; Kim. M. H. 2008, Semijoin-Based Spatial Join Processing in Multiple Sensor Networks, ETRI Journal, 30(6):853-855.
- [6] Kim. M. S; Lee. C. H. 2012, A Middleware System for Efficient Acquisition and Management of Heterogeneous Geosensor Networks Data, Journal of Korea Spatial Information System Society, 20(1):91-103.
- [7] Lee. C. K; Zheng. B; Lee. W; Winter. J. 2007, Materialized In-Network View for Spatial Aggregation Queries in Wireless Sensor Network, ISPRS Journal of Photogrammetry &

- Remote Sensing, 62(5):382-402.
- [8] Meka. A; Singh. A. 2005, DIST: A Distributed Spatio-temporal Index Structure for Sensor Networks, Proc. ACM CIKM, 139-146.
- [9] Park. K; Lee. B; Elmasri. R. 2007, Energy Efficient Spatial Query Processing in Wireless Sensor Networks, Proc. 21st Int'l Conference on Advanced Information Networking and Applications Workshops, 719-724.
- [10] Sharifzadeh. M; Shahabi. C. 2004, Supporting Spatial Aggregation in Sensor Network Databases, Proc. 12th ACM Int'l Workshop on Geographic Information Systems, 166-175.
- [11] Soheili. A; Kalogeraki. V; Gunopulos. D. 2005, Spatial Queries in Sensor Networks, Proc. 14th ACM Int'l Workshop on Geographic Information Systems, 61-70.
- [12] Trigoni. N; Yao. Y; Demers. A; Gehrke. J; Rajaraman. R. 2005, Multi-query Optimization for Sensor Networks, Proc. IEEE DCOSS'05, LNCS 3560, 307-321.
- [13] Xiang. S; Lim. H. B; Tan. K. L; Zhou. Y. 2007, Two-Tier Multiple Query Optimization for Sensor Networks, Proc. ICDCS'07, 39-45.
- [14] Yoon. M; Kim. Y. K; Bista. R; Chang. J. W. 2010, A Data Aggregation Scheme based on Designated Path for Efficient Energy Management of Sensor Nodes in Geosensor Networks, Journal of Korea Spatial Information System Society, 12(1):10-17.

논문접수 : 2013.02.07
수정일 : 2013.04.07
심사완료 : 2013.04.10