

# 센서데이터의 연속적인 스카이라인 질의 처리를 위한 효율적인 필터링기법

(An Efficient Filtering Method for  
Processing Continuous Skyline  
Queries on Sensor Data)

장수민<sup>†</sup>      강광구<sup>†</sup>  
(Su Min Jang)      (Gwang Goo Kang)

유재수<sup>\*\*</sup>  
(Jae Soo Yoo)

**요약** 본 논문은 무선 센서 네트워크의 환경에서 연속적인 스카이라인 질의를 효율적으로 처리하는 새로운 필터링 기법을 제안한다. 기존 필터링기법은 라우터 경로를 기반으로 필터를 생성하고 적용하는 기법이다. 그러나 이러한 기법은 필터의 적용범위가 부분적이기 때문에 센서노드들이 싱크노드에게 전송해야 할 데이터를 대부분 싱크노드로 전송하고 에너지의 효율성 측면에서 좋지 않다. 그래서 본 논문은 적은 비용으로 효율적인 필터를 생성하고 그 필터의 적용범위를 전체 센서노드로 적용하여 싱크노드에 전송해야 할 데이터를 획기적으로 감소시키는 기법을 제안한다. 그 효율적인 필터는 센서데이터의 지역적 특성과 클러스터링을 이용하여 생성한다. 제안된 기법의 우수성을 보이기 위해 다양한 환경에서 시뮬레이션을 통한 성능 평가를 수행한다.

- 이 논문은 2009년 교육과학기술부로부터 지원받아 수행된 연구임(지역거점연구단육성사업/충북BIT연구중심대학육성사업단)
- 이 논문은 2009 한국컴퓨터종합학술대회에서 '센서데이터에 대한 연속적인 스카이라인 질의 처리의 효율적인 필터링기법'의 제목으로 발표된 논문을 확장한 것임

<sup>†</sup> 학생회원 : 충북대학교 정보통신공학과  
jsm@cbnu.ac.kr

kang@netdb.cbnu.ac.kr

<sup>\*\*</sup> 종신회원 : 충북대학교 정보통신공학과 교수  
yjs@cbnu.ac.kr

(Corresponding author)

논문접수 : 2009년 8월 13일

심사완료 : 2009년 10월 12일

Copyright©2009 한국정보과학회 : 개인 목적이나 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. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제15권 제12호(2009.12)

**키워드** : 무선 센서네트워크, 연속적인 스카이라인 질의, 센서 클러스터링, 센서생명주기, 에너지 효율성

**Abstract** In this paper, we propose a novel filtering method for processing continuous skyline queries on wireless sensor network environments. The existing filtering methods use the filter based on router paths. However, because these filters are applied not to a whole area but to a partial area, these methods send almost data of sensor nodes to transmit to the base station and have no sufficient effect in terms of energy efficiency. Therefore, we propose an efficient method to dramatically reduce the transmission data of sensors through applying a low-cost and effective filter to all sensor nodes. The proposed effective filter is generated by using characteristics such as the data locality and the clustering of sensors. An extensive performance study verifies the merits of our new method.

**Key words** : Wireless sensor networks, Continuous skyline queries, Sensor clustering, Sensor life time, Energy efficiency

## 1. 서론

최근 소형기기의 하드웨어 발달 및 무선통신기술의 발달로 다양한 탐지 데이터를 생성하는 무선 센서들이 개발되어 여러 가지 응용분야에서 사용되고 있다. 또한 그와 같은 센서들은 네트워크의 형태로 수백 또는 수천 개의 센서 노드들로 구성되어 이용되고 있다[1,2]. 이러한 무선 센서 네트워크의 환경에서 여러 가지 질의형태 중에 다중 속성을 갖는 데이터에 대한 스카이라인 질의(Skyline query)라는 질의가 크게 대두되고 있다.

스카이라인 질의처리는 Borzonyi[3]에 의해서 소개되었으며, 스카이라인 질의는 전체 객체 집합에서 대상 객체의 여러 속성을 다른 객체가 지배(Dominate)하지 않는 관심 있는 객체 집합을 검색하는 질의이다. 여기서 "지배한다(Dominate)"의 정의는 다음과 같다. 두 개의 객체  $O_1$ 과  $O_2$ 가 있을 경우,  $O_1=(x_1, x_2, \dots, x_d)$ ,  $O_2=(y_1, y_2, \dots, y_d)$ 와 같이  $d$ 개의 속성들을 가지고 있고, 만약  $\forall i, x_i \leq y_i$ , 이고  $\exists j, x_j < y_j$  ( $1 \leq i, j \leq d$ )을 만족한다면  $O_1$ 이  $O_2$ 를 지배한다. 다시 말하면, 모든 속성측면에서  $O_1$ 이  $O_2$ 보다 같거나 하나 이상의 속성에서 우수한 경우  $O_1$ 이  $O_2$ 를 지배하는 것이다.

그럼 1은 질의의 대상이 되는 객체가 호텔들이고 그 호텔과 해변과 사이의 거리와 호텔의 숙박료라는 두 가지 속성을 갖고 있다. 이러한 호텔에 대하여 스카이라인 질의는 "해변에서 가깝고 숙박료가 싼 호텔들을 검색하라." 이다. 이 질의에 대한 적절한 호텔은 여러 개의 호텔들 간에 서로 비교를 통하여 특정호텔에 비하여 해변

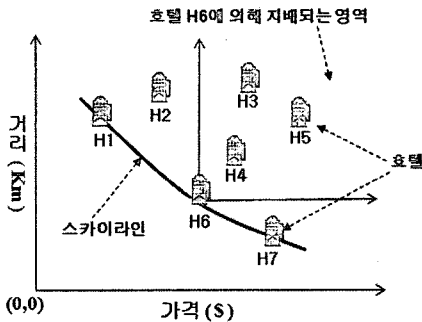


그림 1 스카이라인 질의의 예제

에서 멀면서 숙박료도 비싼 호텔들을 제거한 후 남아있는 호텔들만을 구하는 것이다. 호텔 H6은 거리뿐만 아니라 가격측면에서 호텔 H3, H4, H5를 지배한다.

이처럼 지배되는 호텔들을 모두 제거하면 호텔 H1, H6, H7만이 남는다. 이때 남는 호텔 H1, H6, H7이 스카이라인이다. 위와 같은 방법은 모든 데이터를 수집한 중앙집중식 스카이라인 처리기법이다. 그러나 이러한 처리 기법은 센서네트워크 환경에서는 불필요한 통신이 발생하고, 그에 따라 불필요한 센서의 에너지 소비가 발생한다.

그래서 본 논문에서는 센서 네트워크에서 스카이라인 질의처리를 위해 전송되는 불필요한 통신을 줄여서, 에너지 소비를 줄이는 기법을 제안한다. 제안하는 기법의 특징은 크게 두 가지 필터 적용 범위의 확대와 우수한 필터율을 갖는 필터의 생성이다. 이러한 효율적인 필터링을 통하여 제안하는 기법은 기존의 방법보다 약 30% 이상의 필터율이 높아지는 효과가 있다.

본 논문의 구성은 다음과 같다. 2장에서 관련된 연구를 소개하고 3장에서는 본 논문에서 제안하는 ECF (Excellent Cluster Filter)기법 및 특성을 기술한다. 마지막으로 4장에서 다양한 실험환경에서 제안하는 기법과 MFTAC기법의 성능평가를 통하여 제안하는 필터 기법의 우수성을 입증한다.

## 2. 관련연구

스카이라인 질의에 대한 기존 기법들 중에 대표적인 것으로 BNL(Block Nested Loop)기법[4]과 R-tree기법에서 NN(Nearest Neighbor)을 이용한 BBS(Branch and Bound Skyline)기법[5]가 제안 되었다. 위와 같은 기법들은 모든 측정값을 베이스 스테이션으로 보낸 후 베이스 스테이션 측에서 질의를 처리하는 중앙집중식 질의 처리 기법들이다. 이러한 중앙집중식 질의처리 기법들은 이벤트가 빈번하게 발생하는 센서 네트워크에는 에너지 효율성이 떨어진다. 모든 센서들은 파워가 제한적이기 때문에 에너지 효율적인 질의처리가 필요하다[6].

센서 네트워크 환경에서 기존의 스카이라인 처리 기법인 MFTAC기법(MFT applied Aggregation method based on Cluster)[7]은 전체 센서 노드에서 싱크노드로 전송해야 할 데이터를 필터링할 수 있는 MFT(Min Score Filtering Tuple)필터를 배포하여 질의 결과에 포함 되지 않는 불필요한 데이터의 전송을 차단한다. 하지만 MFTAC기법은 연속질의에 적용하게 되면 매 라운드 마다 필터를 배포하기 위한 과도한 에너지 소모를 발생시킨다. 또한 MFTAC기법은 필터를 설정하기 위해서 라우팅 트리의 상위 노드로부터 최적의 MFT를 계산하고, 하위 노드로 배포한다. 각 센서 노드는 상위 노드에 의해 결정된 MFT를 이용하여 필터링을 수행한다. 하지만, MFT는 전송 경로상의 부분 영역에서만 반영하기 때문에 많은 False Positive 데이터가 보고되고, 이것은 네트워크 수명을 단축시키는 주된 원인이 된다.

MFTAC기법은 부분적인 영역에 해당하는 센서들의 데이터를 서로 비교하여 필터를 생성하고 그 필터의 적용범위도 한정되기 때문에 에너지 효율적인 질의처리가 되지 않는 문제점을 갖고 있다. 그림 2는 MFTAC기법과 같은 지역적 필터링을 이용한 연속적인 스카이라인 질의 처리를 보여준다. 위쪽에서 질의를 보내면서 아래쪽으로 필터링하는 Top-Down방식이나 그 반대인 Bottom-Up 방식은 라우터 경로에 따라 필터를 갱신하고 필터를 적용하는 방식이다. 지역적 필터 LF4나 LF6은 스카이라인 질의에서 스카이라인에 해당하는 데이터 값을 갖는 센서들의 영역에서 생성된 필터이다. 그러나 필터 LF4나 LF6에 적용되는 센서들은 지역적 특성으로 비슷한 값을 갖기 때문에, 그 생성된 필터에 의해서 주변 센서들이 싱크노드(BS)로 전송해야 할 데이터가 효과적으로 감소되지 않는 문제점을 갖고 있다. 그러나 그 필터 LF4나 LF6을 센서 S1부터 S6로 구성되어 있는 라우터 경로에 적용하였다면 필터링 효과가 매우 클 것이다.

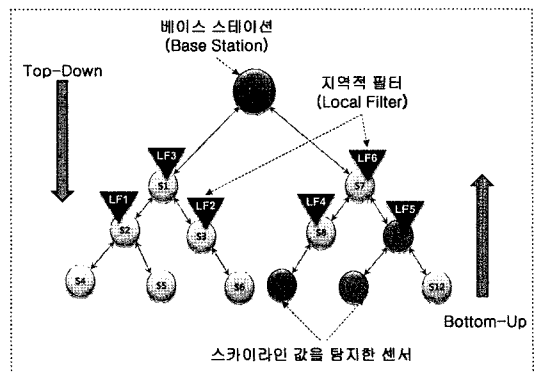


그림 2 지역적 필터링을 이용한 연속적인 스카이라인 질의 처리

### 3. 제안하는 ECF(Excellent Cluster Filter)기법

본 논문은 스카이라인 질의처리에서 많은 False Positive 데이터가 전송됨으로써 네트워크 수명을 단축시키는 문제점을 해결하기 위하여 에너지 효율적인 질의처리 기법을 제안한다. 제안하는 ECF기법은 우선 전체적인 센서들의 데이터를 비교하여 우수 클러스터를 선정하고 우수 클러스터에 해당하는 일부분의 센서들만 탐지하여 최적화된 필터를 생성한다. 그리고 그 필터를 전체 영역의 센서들에게 적용함으로써 불필요한 센서 데이터가 싱크노드(BS)에 전송되는 것을 최소화한다. 실제 센서데이터들은 일정한 시간과 지역적으로 유사한 데이터를 발생시키는 성향을 갖기 때문에 우수 클러스터에서 생성된 필터의 사용은 부분적인 영역에서 생성된 필터를 사용한 것보다 필터링 효과가 뛰어나다. 제안하는 ECF기법의 연속적인 스카이라인질의 처리절차는 아래와 같다.

1. 전체 센서노드로부터 모든 데이터를 전송받아서 클러스터링과 우수 클러스터를 선정한다.
2. 우수 클러스터 헤더에 질의를 배포하고 그 클러스터에 포함된 센서들의 데이터를 조합하여 ECF(Excellent Cluster Filter)를 생성한다.
3. 생성된 우수 필터와 질의를 클러스터들에게 배포하고 그 우수 필터를 이용한 필터링을 통하여 필요한 데이터만을 싱크노드로 전송하여 질의를 처리한다.
4. 우수 클러스터의 재선정이 필요 없는 경우 2번 과정과 3번 과정을 반복한다. 우수 클러스터의 재선정이 필요할 경우 1번 과정부터 시작한다.

우수 클러스터의 선정기준은 센서데이터의 값의 스카이라인 질의에서 지배하는 영역의 크기의 합으로 선정한다. 총 지배영역의 크기가 가장 큰 클러스터를 우수 클러스터로 선정한다. 지배영역이 크다는 것은 그만큼 필터링할 수 있는 데이터의 범위가 크기 때문에 데이터의 전송량을 감소시키는 확률이 높다. 그림 3은 ECF의 생성 및 필터링 과정을 보여준다. 3개의 클러스터 C1, C2, C3가 있고 그중에 클러스터 C2가 스카이라인에 해당하는 값을 갖는 센서(S8, S9, S10) 3개가 포함되어 있다. 꼭 스카이라인에 해당하는 센서노드 수와 지배영역의 크기가 비례적이지 않지만, 스카이라인에 해당하는 센서노드 수가 많은 곳이 큰 지배영역을 가질 확률이 높다. 그래서 스카이라인에 해당하는 센서노드 수가 많은 C2가 우수 클러스터로 선정되고, 그 클러스터의 헤더 CH2가 센서 S8, S9, S10의 데이터 값을 조합하여 ECF를 생성한다. 그리고 그 ECF를 우수 클러스터 C2를 제외한 C1, C3에 전송하여 ECF에 의한 필터링을 한다.

본 논문에서 제안하는 ECF기법이 센서들의 에너지 효율성을 극대화하기 위해서는 우수 클러스터를 효율적

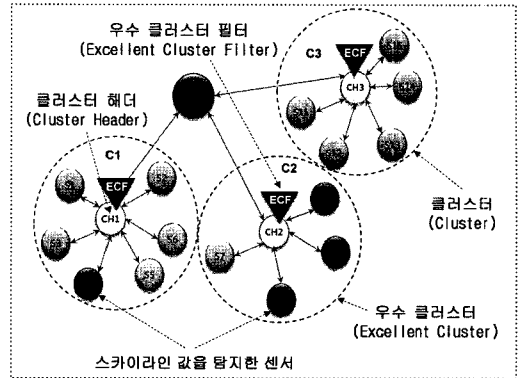


그림 3 ECF의 생성 및 필터링 과정

으로 선정하는 것이 중요하다. 일반적으로 센서의 데이터 값은 지역적으로 유사한 값을 갖는 특성이 있지만 일정한 시간흐름에 따라 또는 데이터의 특성에 따라 그 값들이 전체적으로 변화하게 된다. 이처럼 센서들의 데이터 값이 전체적으로 변화가 생긴다면 우수 클러스터를 재선정할 필요가 있다. 그래서 우수 클러스터의 재선정을 위한 시점을 결정하기 위해서 본 논문에서는 우수 클러스터에서 생성된 ECF가 전체 센서들이 전송해야 할 데이터에 대한 필터율(FRATE)을 K퍼센트보다 낮아지는 시점이나 일정한 M시간 주기를 우수 클러스터의 재선정 시점으로 결정한다(K, M은 사용자가 정한 상수). FRATE은 ECF의 필터링을 통하여 실제적으로 전송된 데이터량(TDATA(ECF))와 필터를 사용하지 않고 전송되어야 할 전체 데이터량(TDATA)을 이용하여 아래와 같은 식 (1)으로 산출한다.

$$FRATE = TDATA(ECF) / TDATA \quad (1)$$

그림 4는 제안하는 ECF기법에 의한 필터링 절차과정을 보여준다. 이처럼 제안하는 기법은 전체영역에 대한 필터 ECF의 필터링과 ECF필터링에도 남아 있는 센서 데이터를 다시 한 번 지역적 필터링으로 실질적으로 BS에 전송되는 센서데이터를 획기적으로 감소시킨다. 이처럼 전송할 센서데이터의 감소는 센서의 제한된 에너지 측면에서 효율적이다.

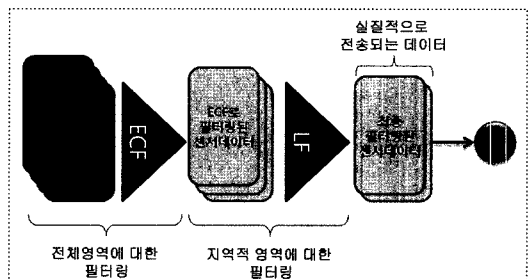


그림 4 제안하는 ECF기법에 의한 필터링 절차과정

4. 성능평가

본 논문의 성능평가를 위하여 시뮬레이션을 통하여 MFTAC기법과 제안하는 ECF기법을 비교하였다. 시뮬레이션 환경은 C언어를 이용하여 구축하였으며, 100개의 센서 노드를 100m×100m 크기의 평지에 균일하게 배치했고 BS를 (50m, 50m)에 위치시켰다. 각 센서 노드의 통신 가능거리는 30m로 설정하였다. 각 센서에서 발생하는 데이터 값은 실 데이터 값과 가까운 real\_data형과 일정범위의 값을 랜덤하게 생성되는 random\_data형으로 설정하였다. real\_data형은 기상청에서 2000년부터 2008년까지의 수집한 실 센서데이터를 바탕으로 생성하였다. 그림 5는 안면도에서 측정한 자외선B와 이산화탄소 화탄소의 기후데이터를 이용하여 생성된 데이터이다.

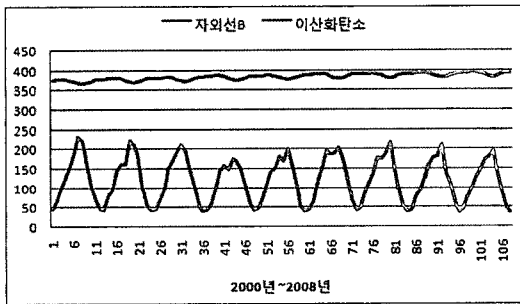


그림 5 안면도의 기후데이터(real\_data형)

random\_data형의 데이터는 0에서 100사이의 정수 값을 갖도록 설정하였다. 성능평가는 전체 전송되어야 할 데이터 중 제거된 데이터의 비율을 측정함으로써 센서에 에너지의 효율성을 측정하였다. 그림 6은 real\_data형의 데이터에 대한 제안하는 ECF기법과 MFTAC기법의 필터율(FRATE)을 비교하였다. 제안하는 기법이 MFTAC기법에 비교하여 약 5%~30%이상 필터율이 향상되는 것을 보여준다. 이는 ECF를 센서의 전체영역에 적용하여 센서데이터의 전송률을 줄이는 효과가 있기 때문이다.

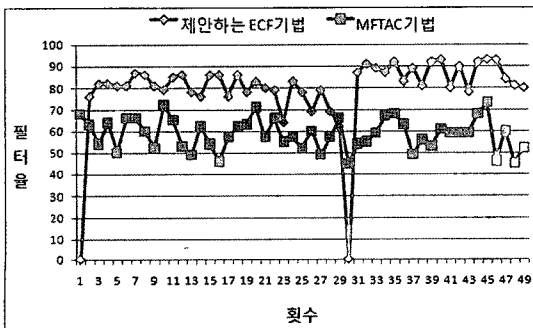


그림 6 real\_data형의 데이터에 대한 필터율

제안하는 ECF기법은 30회마다 우수 클러스터를 재정정하기 위하여 전체 센서데이터를 모두 전송받기 때문에 필터율이 0%이다. 그림 7은 random\_data형의 데이터에 대한 성능평가이다. 제안하는 ECF기법의 필터율이 MFTAC기법에 비하여 약 5%에서 15%정도 향상되는 것을 보여준다.

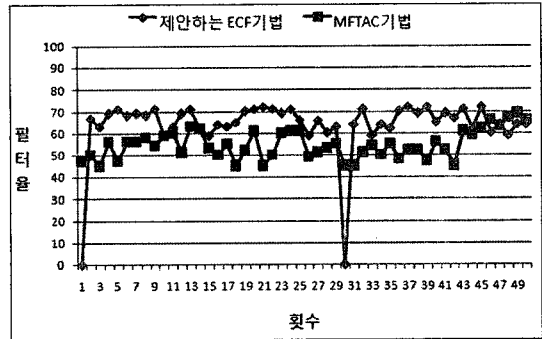


그림 7 random\_data형의 데이터에 대한 필터율

random\_data형이 real\_data형보다는 센서의 데이터 값의 변화가 심하기 때문에 real\_data형의 필터율보다 다소 낮다. 그 이유는 센서의 데이터 값의 지역적 특성을 반영하지 못하기 때문에 우수 클러스터에서 생성된 필터가 real\_data형만큼 효과적이지 못하기 때문이다. 그래도 MFTAC기법의 라우팅 경로에 따른 제한된 영역의 필터링보다는 제안하는 ECF기법이 높은 필터율을 보여준다.

5. 결론

본 논문은 우수 클러스터를 생성하여 연속적인 스카이라인 질의를 처리하기 위한 효율적인 필터링 기법을 제안하였다. 우수 클러스터에서 생성된 ECF 필터를 전체 센서에 적용함으로써 전송데이터를 획기적으로 감소시켰다. 향후연구방향은 일반적인 클러스터링기법이 아닌 센서들의 에너지측면과 데이터 특성을 고려한 최적화된 클러스터링 알고리즘과 클러스터 크기에 따른 최적화 과정이 필요하다.

참고문헌

[1] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," In *Proceedings of the ACM Mobile Computing and Networking*, pp.56-67, 2000.

[2] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a Tiny Aggregation Service for Ad-Hoc Sensor Networks," In *Proceedings of the*

*Operating Systems Design and Implementation*, pp.131-146, 2002.

- [3] S. Borzsonyi, D. Kossmann, and K. Stocker, "The skyline operator," In *Proceedings of the International Conference on Data Engineering*, pp.421-430, 2001.
- [4] J. Chomicki, P. Godfrey, J. Gryz, and D. Liang, "Skyline with presorting," In *Proceedings of the International Conference on Data Engineering*, pp.717-719, 2003.
- [5] Dimitris Papadis, Yufei Tao, Greg Fu, Bernhard Seeger, "An Optimal and Progressive Algorithm for Skyline Queries," In *Proceedings of the ACM SIGMOD*, pp.467-478, 2003.
- [6] A. Sharaf, J. Beaver, A. Labrinidis, and K. Chrysanthis, "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks," *VLDB Journal*, vol.13, no.4, pp.384-403, 2004.
- [7] Y Kwon, JH Choi, YD Chung, S Lee, "In-Network Processing for Skyline Queries in Sensor Networks," *IEICE Transactions on Communications*, vol.E90-B, no.12, pp.3452-3459, 2007.