

# 공간 분할 방법을 사용하는 Skyline 질의 방법의 문제점 비교 분석

임선영\*, 박은영\*\*, 박영호\*

\*숙명여자대학교 멀티미디어학과

\*\*협성대학교 시각디자인학과

e-mail: sunnyihm@sm.ac.kr, parkey@uhs.ac.kr, yhpark@sm.ac.kr

## A Problem Analysis of Skyline Queries using Space Partitioning Techniques

Sun-Young Ihm\*, Eun-Young Park\*\*, Young-Ho Park\*

\*Dept of Multimedia Science, Sookmyung Women's University

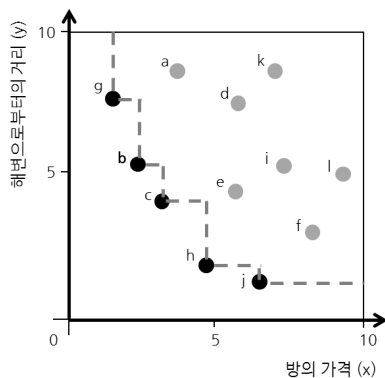
\*\*Dept of Visual Design, HyupSung University

### 요 약

최근 대용량 혹은 복잡한 데이터에서의 빠른 검색을 돕는 Skyline 질의 처리에 대한 관심이 높아지고 있으며, 많은 응용프로그램에서 사용되고 있다. Skyline 질의는 데이터베이스의 튜플들을 공간상으로 표현하여 질의 처리를 할 수 있는 데, 이 때 빠른 처리를 위하여 공간 분할 기법이 사용된다. 본 논문에서는 공간 분할 기법을 사용한 Skyline 질의 방법들을 소개하고 문제점을 분석한다. 또한, 문제점 해결 방안을 함께 제시하며 기대 효과를 예측한다.

### 1. 서론

Skyline 질의는 많은 응용프로그램에서 사용되어왔으며 질의 처리에 대한 많은 연구가 이루어지고 있다. Skyline 질의는 사용자들이 원하는 데이터를 복잡하거나 대용량의 데이터에서 빠르게 검색할 수 있도록 도와준다. 데이터베이스에서 호텔을 검색하는 경우를 예를 들어 설명하면, 데이터베이스의 각 튜플들은 공간에서 하나의 오브젝트로 표현될 수 있다. 또한 각 오브젝트의 속성들은 차원으로 표현이 가능하다. 예를 들어 각 호텔은 '방의 가격'과 '해변으로부터의 거리'라는 속성을 가질 수 있으며, '방의 가격'은 x축으로, '해변으로부터의 거리'는 y축으로 표현이 가능하다. 그림 1은 본 예제를 나타내고 있다. 이 때, Skyline을 이용하여 레이어 기반의 인덱스를 구축해 놓으면 사용자가 원하는 호텔을 쉽고 빠르게 검색할 수 있다.



(그림 1) Skyline 질의의 예

공간으로 표현된 데이터베이스에서 Skyline을 이용하여 인덱스를 빠르게 구축하기 위한 방법으로 공간 분할 기법이 다양하게 연구되어 왔다. 본 논문에서는 공간 분할 기법을 사용한 Skyline 질의 처리 방법들을 조사하고 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 대표적인 연구들을 설명한다. 3장에서는 기존 방법의 문제점을 비교 분석하고 이를 해결하기 위한 방법에 대하여 간략하게 논의한다. 마지막으로 4장에서는 기대효과와 함께 결론을 내린다.

### 2. 기존 연구

본 장에서는 Skyline 질의 처리 방법 중 공간 분할 기법을 사용하는 연구들을 소개하며 장점 및 단점을 설명하고 문제점을 분석한다. 공간 분할 기법을 사용하는 Skyline 질의 처리 방법의 목적은 데이터 공간을 분할하여 처리하여 더 많은 오브젝트들을 배제하여 계산 시간을 줄이는 것이다. 또한 공간을 분할하게 되면 각 분할된 공간을 서로 다른 서버에서 동작하게 하여 처리 속도를 향상시킬 수도 있다. 공간을 분할하여 Skyline 계산을 한 후에는 다시 합치는 작업이 필요하다.

이러한 공간 분할 방법을 사용하는 Skyline 질의 처리 방법 중 대표적인 연구는 Grid 기반의 공간 분할 기법을 사용하는 Skyline 질의, Angle기반의 공간 분할 기법을 사용하는 Skyline 질의[1] 그리고 Plane-Projection 기반의 공간 분할 기법을 사용하는 Skyline 질의[3] 연구가 있다. 또한, 이 외에도 [2, 4, 5, 6] 등에서도 연구가 진행되었다.

### 2.1 Grid 기반의 공간 분할 기법을 사용하는 Skyline 질의 (GPS)

Grid 기반의 공간 분할 기법은 공간을 분할하는 방법 중 가장 단순하며 효과적인 방법이다. 공간을 격자 형태로 분할하여 각 분할된 공간에서 Skyline을 구하는 방법이다. 분할 방법이 매우 간단하다는 장점이 있어서 분산 환경 등 Skyline 질의를 사용하는 많은 응용 프로그램에서 사용되고 있다.

### 2.2 Angle 기반의 공간 분할 기법을 사용하는 Skyline 질의 (APS)

Angle 기반의 공간 분할 기법은 GPS를 개선한 방법으로 각도를 이용하여 데이터 공간을 분할하여 Grid 기반의 공간 분할 방법에 비하여 각 분할된 공간에서 더 적은 수의 Skyline 오브젝트를 얻을 수 있다. 결과적으로 분할된 Skyline 오브젝트들을 합치는 과정에서 계산이 줄기 때문에 성능이 좋다.

### 2.3 Plane-Projection 기반의 공간 분할 기법을 사용하는 Skyline 질의 (PPPS)

Plane-Projection 기반의 공간 분할 기법은 APS를 개선한 방법으로 데이터 공간을 분할하기 위하여 프로젝션을 사용한다. 각 튜플의 값들을 hyperplane에 프로젝션하여 공간을 분할한다. 예를 들어 튜플이 3개의 속성을 가지고 있다면 각 속성 값을  $x + y + z = 1$  이라는 hyperplane에 프로젝션 한다.  $d$ 개의 속성을 가지는 튜플의 각 값에 대하여 프로젝션 값을 구하는 식은 다음과 같다.

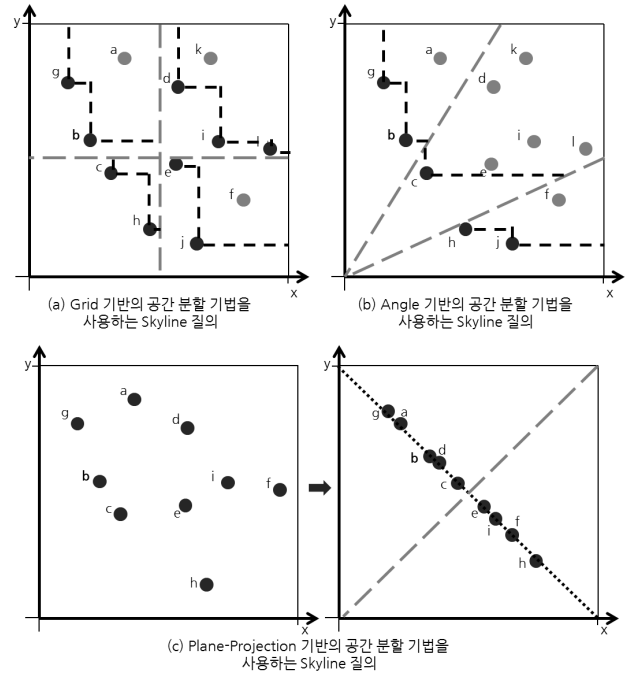
$$(x_1, \dots, x_d) \Rightarrow (x_1, \dots, x_d) \times \frac{1}{x_1 + \dots + x_d}$$

### 3. 문제점 비교 분석

본 장에서는 2장에서 설명한 기존 연구의 문제점을 비교 분석하고, 해결방안을 간략하게 소개한다.

GPS의 문제점은 분할된 공간 중 다수가 전체 결과와 관계가 없는 경우가 많기 때문에 특히 병렬 처리를 이용하여 Skyline 질의 처리를 하는 경우에는 성능이 더 좋지 않다. APS는 결과 집합에 포함되지 않는 불필요한 오브젝트들에 대한 불필요한 접근이 많다는 단점이 있다. PPPS는 공간을 분할하기 전에 근사 Skyline을 구하는 전처리 단계를 적용하여 성능을 더욱 향상 시켰지만 여전히 불필요한 오브젝트에 대한 불필요한 접근이 발생한다는 문제점이 있다. GPS, APS, PPPS가 Skyline 질의를 하는 방법을 그림 2 에서 비교하였다. 그림 2 (a), (b), (c)는 각각 GPS와 APS, PPPS를 표현하고 있다.

이러한 문제점을 해결하기 위해서는 분할된 공간에서의 Skyline 질의 처리 시 성능을 높이기 위하여 효과적으로 분할하거나 더 많은 오브젝트들을 배제해야 한다. 분할된 공간에서 계산된 Skyline 오브젝트의 수가 적으면 마지막에 병합하는 과정에서의 계산이 줄기 때문이다.



(그림 2) 공간 분할 기법을 사용하는 Skyline 질의 비교

### 4. 결론 및 기대효과

본 논문에서는 기존의 공간 분할 기법을 사용하는 Skyline 질의 처리 연구들을 설명하고 문제점을 분석하였다. 기존의 방법들은 분할된 공간에서의 Skyline 계산을 줄이기 위하여 Grid 기반, Angle 기반 그리고 Plane-Projection 기반의 공간 분할 기법을 사용하였다. 하지만 여전히 분할된 공간에서 Skyline 질의 처리를 할 때 결과가 아닌 오브젝트에 불필요한 접근이 발생한다. 이에 대한 해결책으로는 더 많은 오브젝트를 배제해야 한다. 결과적으로 Skyline 질의 처리 성능이 우수해 질 것을 기대해 볼 수 있다.

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2012-003797)

### 참고문헌

[1] A. Vlachou, C. Doukeridis, Y. Kotidis "Angle-based Space Partitioning for Efficient Parallel Skyline Computation" Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp.227-238, 2008  
 [2] B. Cui, H. Lu, Q. Xu, L. Chen, Y. Dai, Y. Zhou "Parallel distributed processing of constrained skyline queries by filtering" In ICDE, pp.546-555, 2008  
 [3] H. Kohler, J. Yang, X. Zhou "Efficient Parallel Skyline Processing using Hyperplane Projections" Proceedings of the 2011 ACM SIGMOD International

Conference on Management of data, pp.85-96, 2011

[4] P. Wu, C. Zhang, Y. Feng, B. Y. Zhao, D. Agrawal, A. E. Abbadi "Parallelizing skyline queries for scalable distribution" In EDBT, pp.112-130, 2006

[5] S. Ratnasamy, P. Francis, M. Handley, R. Karp, S. Shenker "A scalable content-addressable network" In SIGCOMM, pp.161-172, 2001

[6] S. Wang, B. C. Ooi, A. K. H. Tung "Efficient skyline query processing on peer-to-peer networks" In ICDE, pp.1126-1135, 2007