

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2009-0100852 (43) 공개일자 2009년09월24일
(51) Int. Cl. <i>G06F 17/30</i> (2006.01) (21) 출원번호 10-2008-0026332 (22) 출원일자 2008년03월21일 심사청구일자 2008년03월21일	(71) 출원인 고려대학교 산학협력단 서울 성북구 안암동5가1 고려대학교 내 (72) 발명자 이상근 서울 동대문구 제기2동 153 안암골 벽산아파트 101동 203호 최현식 서울 성북구 안암동5가 102-100번지 301호 구원교 서울 노원구 공릉동 81번지 태강아파트 1005동 602호 (74) 대리인 현중철	

전체 청구항 수 : 총 20 항

(54) 셀 상태 값을 이용한 연속 스카이라인 질의 방법, 그 시스템 및 이를 기록한 기록매체

(57) 요약

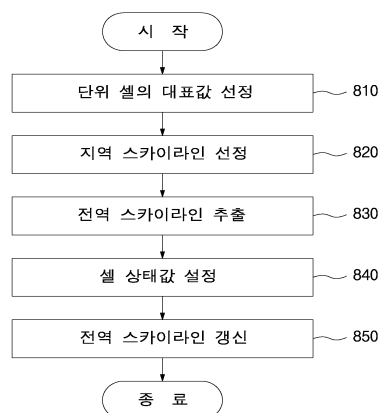
셀 상태 값을 이용한 연속 스카이라인 질의 방법, 그 시스템 및 이를 기록한 기록매체가 개시된다.

본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 방법은,

다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자 형태의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정하는 단계; 상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정하는 단계; 상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출하는 단계; 상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정하는 단계; 및 상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신하는 단계를 포함한다.

본 발명에 의하면, 튜플들의 속성이 빈번하게 바뀌는 환경에서도 스카이라인 결과값을 항상 최신의 값으로 유지할 수 있고, 셀 방문에 있어서의 중복 방문횟수를 줄임으로써 시스템의 성능을 향상시킬 수 있으며, 셀 방문에 따른 유지 비용을 감소시킬 수 있는 효과가 있다.

대표도 - 도8



특허청구의 범위

청구항 1

다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정하는 단계;

상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정하는 단계;

상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출하는 단계;

상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정하는 단계; 및

상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신하는 단계를 포함하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 2

제 1 항에 있어서,

상기 단위 셀의 대표값은

원점과 가장 가까운 거리에 대응되는 셀의 좌표이며, 상기 단위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현되는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 3

제 1 항에 있어서,

상기 지역 스카이라인을 선정하는 단계는

상기 단위 셀 내에 포함되는 튜플 들의 다중 속성값을 비교하는 단계; 및

상기 단위 셀 내에 포함되는 튜플들 중에서 하나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정하는 단계를 포함하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 4

제 3 항에 있어서,

상기 전역 스카이라인을 선정하는 단계는

상기 선정된 지역 스카이라인의 다중 속성값을 비교하는 단계; 및

상기 지역 스카이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정하는 단계를 포함하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 5

제 1 항에 있어서,

상기 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역이고,

상기 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 지배 영역을 감싸는 셀들로 이루어진 영역이고,

상기 반 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이

작은 대표값을 가지는 셀들로 이루어진 영역이고,

상기 반 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역인 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 6

제 5 항에 있어서,

상기 각 셀에 대한 셀 상태값을 설정하는 단계는

상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 *SK* 상태로 설정하고,

상기 단위 셀이 상기 지배 영역에 해당하면 *DOM* 상태로 설정하며,

상기 단위 셀이 상기 *SK* 및 *DOM*의 어느 상태에도 속하지 않으면 *FREE* 상태로 설정하는 단계인 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 7

제 6 항에 있어서,

상기 전역 스카이라인을 갱신하는 단계는

상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 *DOM*이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하는 단계;

상기 추가되는 튜플이 포함되는 셀의 상태값이 *FREE* 또는 *SK* 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하는 단계; 및

상기 추가되는 튜플이 포함되는 셀의 상태값이 *FREE* 또는 *SK* 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정하는 단계를 포함하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 8

제 6 항에 있어서,

상기 전역 스카이라인을 갱신하는 단계는

상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 *DOM*이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하는 단계;

상기 삭제되는 튜플이 포함되는 셀의 상태값이 *SK* 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 띠 및 지배 띠를 방문하여 전역 스카이라인의 여부를 결정하는 단계; 및

상기 삭제되는 튜플이 포함되는 셀의 상태값이 *SK* 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플만의 지배 영역 및 지배 띠를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정하는 단계를 포함하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 9

제 1 항에 있어서,

상기 단위 셀 C_{ij} 은

상기 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 상기 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 을 저장하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 10

제 1 항에 있어서,

상기 전역 스카이라인을 갱신하는 단계는

상기 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인의 전역 스카이라인의 튜플을 사용자에게 의해 설정된 주기마다 출력하는 단계를 더 포함하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 방법.

청구항 11

제 1 항 내지 제 10 항 중 어느 한 항의 방법을 컴퓨터에서 수행시킬 수 있는 프로그램으로 기록된 기록매체.

청구항 12

다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정하는 셀 대표값 선정부;

상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정하는 지역 스카이라인 선정부;

상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출하는 전역 스카이라인 추출부;

상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정하는 셀 상태값 설정부; 및

상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신하는 스카이라인 갱신부를 포함하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 13

제 12 항에 있어서,

상기 단위 셀의 대표값은

원점과 가장 가까운 거리에 대응되는 셀의 좌표이며, 상기 단위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현되는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 14

제 12 항에 있어서,

상기 지역 스카이라인 선정부는

상기 단위 셀 내에 포함되는 튜플 들의 다중 속성값을 비교하고, 상기 단위 셀 내에 포함되는 튜플들 중에서 하

나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 15

제 14 항에 있어서,

상기 전역 스카이라인 추출부는

상기 선정된 지역 스카이라인의 다중 속성값을 비교하고, 상기 지역 스카이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 16

제 12 항에 있어서,

상기 셀 상태값 설정부는

상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역을 지배 영역으로 설정하고,

상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 지배 영역을 감싸는 셀들로 이루어진 영역을 지배 띠로 설정하며,

상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 작은 대표값을 가지는 셀들로 이루어진 영역을 반 지배 영역으로 설정하며,

상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역을 반 지배 띠로 선정하여 셀 상태 값을 설정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 17

제 16 항에 있어서,

상기 셀 상태값 설정부는

상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 SK 상태로 설정하고, 상기 단위 셀이 상기 지배 영역에 해당하면 DOM 상태로 설정하며, 상기 단위 셀이 상기 SK 및 DOM 의 어느 상태에도 속하지 않으면 $FREE$ 상태로 설정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 18

제 17 항에 있어서,

상기 스카이라인 갱신부는

상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하고,

상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하며,

상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추

가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 19

제 17 항에 있어서,

상기 스카이라인 갱신부는

상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 DOM_i 이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하고,

상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 띠 및 지배 띠를 방문하여 전역 스카이라인의 여부를 결정하며,

상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플 만의 지배 영역 및 지배 띠를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

청구항 20

제 12 항에 있어서,

상기 단위 셀 C_{ij} 은

상기 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 상기 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 을 저장하는 것을 특징으로 하는 셀 상태 값을 이용한 연속 스카이라인 질의 시스템.

명세서

발명의 상세한 설명

기술 분야

<1> 본 발명은 데이터베이스 시스템에서의 질의 처리에 관한 것으로서, 특히 튜플들의 속성이 빈번하게 바뀌는 동적인 환경에서도 스카이라인 결과값을 항상 최신의 값으로 갱신할 수 있고, 스카이라인 결과값 추출에 대한 연산량을 줄임으로써 데이터베이스 처리에 있어 연산 비용을 감소시킬 수 있는 셀 상태 값을 이용한 연속 스카이라인 질의 방법, 그 시스템 및 이를 기록한 기록매체에 관한 것이다.

배경 기술

<2> 스카이라인 질의(Skyline Query)는 다중 속성을 가지는 튜플(Tuple)들의 집합에서 다른 어떤 튜플들 에게도 지배(Domination)되지 않는 튜플을 찾는 문제이다.

<3> 스카이라인에서 지배의 의미는 중요한데 지배의 정의는 하기와 같다. 튜플 r 과 r' 가 있고, r 과 r' 의 모든 속성을 각각 비교하였을 때, r 의 모든 속성들이 r' 의 모든 속성들보다 작거나 같고, r 이 r' 보다 최소한 한 속성이 반드시 작다면 r 이 r' 를 지배한다고 할 수 있다.

<4> 이는 다른 분야에서, 응용값에 따라 큰 것이 작은 값을 지배한다고 할 수도 있으나, 스카이라인 질의에서는 작

은 값이 큰 값을 지배한다고 정의한다. 이는 하기의 내용으로 더욱 명확하게 이해할 수 있다.

- <5> 도 1은 스카이라인 질의에 적용되는 속성값 비교 그래프를 도시한 것이다.
- <6> 전술한 스카이라인 질의를 설명하기 위하여 상기의 도 1과 같이 호텔의 두 가지 속성인 '투숙비'와 '해변으로부터의 거리'를 x, y 좌표값으로 하여 2차원 좌표평면 위의 점인 a, b, c, d, e, f, g 로 표현하기로 한다.
- <7> 이때, 호텔 a 는 해변으로부터의 거리와 투숙비 모두 호텔 f 보다 작기 때문에 a 가 f 를 지배한다고 한다.
- <8> 여기서, c 또한 e, f, g 보다 모든 속성값이 작기 때문에 e, f, g 를 지배한다.
- <9> 그러나, b 와 d 의 관계는 이와는 다르다. b 가 d 보다 거리는 짧지만 투숙비는 같다. 이런 경우에도 스카이라인 질의에는 앞서 정의한 대로 b 가 d 를 지배한다고 할 수 있다.
- <10> 반면, a 와 b 의 경우에는 a 가 b 보다 거리는 짧고, b 가 a 보다 투숙비는 싸다. 이런 경우에는 비교불가(incomparable)라고 한다.
- <11> 결과적으로는, 상기 도 1의 튜플 중에서 스카이라인 튜플은 어떠한 점에게도 지배되지 않는 a, b, c 가 된다.
- <12> 이와 같은 특징을 가지는 스카이라인은 다중 속성을 고려한 의미로써, 데이터 베이스 분야에서 집중적으로 연구되어 왔다.
- <13> 그러나, 기존 스카이라인 연구들은 최근 발전하는 모바일(Mobile) 및 스트림 데이터(Stream data)와 같이 데이터가 실시간으로 변하는 동적인 환경에는 적합하지 않다.
- <14> 기존 정적 환경에서는 한 묶음의 정적 데이터에 대한 한번의 질의 처리를 요구하는데 반하여 동적인 환경은 데이터에 대한 계속적인 질의처리를 요구한다.
- <15> 따라서, 정적 데이터에 대한 질의만을 고려했던 기존 연구들은 데이터가 실시간으로 변하는 동적인 환경에서는 적합하지 않다.
- <16> 이와 같은 문제를 해결하기 위하여, Tao et al.은 스트림 환경에서의 연속적인 스카이라인 질의처리 기법을 제안하였다.
- <17> 그러나, R(Rectangular)-Tree를 기반으로 한 색인을 사용하기 때문에 빈번한 데이터 추가, 삭제 및 데이터 값 변경시의 색인 갱신에 대한 가비용 문제가 발생한다.
- <18> 또한, 매번 튜플 추가시, 해당 튜플의 스카이라인 여부를 판단하기 위하여 범위 질의(Range Query) 수행으로 인한 과비용 문제가 유발된다.
- <19> Li Tian et al.은 상기의 문제를 해결하기 위하여 격자 셀을 색인으로 하는 연속적인 스카이라인 질의 처리(Continuous Monitoring Skyline Query:CMSQ)를 제안하였다.
- <20> 여기서, 연속적인 스카이라인 질의 처리는 영향 영역(Influence Region)에 삽입된 튜플에 한해서만 추가적인 연산을 수행함으로써 연산 비용을 감소시켰다.
- <21> 그러나, 이는 상기 영향 영역을 연결 리스트(Linked List)로 유지하기 위하여 매번 튜플이 결과에 영향을 주는지의 여부를 확인하기 위해서는 $O(n)$ 의 비용이 든다.
- <22> 그리고, 어떤 튜플의 스카이라인 여부를 알기 위하여 시스템이 유지하는 모든 스카이라인과 순차적으로 비교하므로, 큰 비용이 발생하게 된다.
- <23> 더군다나, 차원이 높아짐에 따라 스카이라인의 수는 더욱 증가하기 때문에 차원 증가에 따라 더 큰 비용이 요구된다.
- <24> 한편, 셀 방문시의 큐(Queue)에 현재 방문 셀의 인접 셀들을 넣어 하나씩 꺼내면서 방문하는데, 이때 고차원에서 큐의 크기가 너무 증가한다.

- <25> 또한, 인접 셀 방문 알고리즘이 가지는 특성 때문에 중복 방문 문제가 발생하므로 중복 방문을 회피하기 위한 추가 비용이 더 요구된다.
- <26> 이와 같이, 현재까지의 스카이라인 연구는 크게 정적 데이터를 기반으로 한 정적 스카이라인 질의처리와 지속적으로 변하는 데이터를 기반으로 하는 연속적인 스카이라인 질의처리로 구분할 수 있다.
- <27> 대표적인 정적인 스카이라인 질의 처리 기법은 Block-Nested-Loops(BNL), Divide and Conquer(DC), 최근접(Nearest Neighbor) 질의 기반의 스카이라인, 및 Branch and Bound Search(BBS)가 있다.
- <28> 또한, 연속적인 스카이라인 질의처리 기법은 전술한 Tao et al.의 스트림 환경에서 슬라이딩 윈도우를 사용한 스카이라인 질의와 Li Tian et al.의 격자 색인을 이용한 기법이 있다.
- <29> 우선, Block-Nested-Loops(BNL), Divide and Conquer(DC)는 Borzsonyi et al.에 의해 최초로 제안되었으며, 이것이 최초의 스카이라인 연구이다.
- <30> 그 중 BNL은 가장 원시적인(naive) 알고리즘으로써 우선 첫 번째 튜플을 스카이라인 후보 리스트에 넣고, 이 리스트의 튜플들과 나머지 튜플들을 순차적으로 비교한다.
- <31> 비교 과정에서 후보 리스트에 있는 스카이라인은 후보 리스트에서 제거된다. 이는 가장 직관적이고 적용할 수 있는 응용이 많은 반면 후보 리스트를 메모리에 유지하는 비용이 매우 크다.
- <32> 최근접(Nearest Neighbor:NN) 기반의 스카이라인 질의는 Kossman et al.에 의해 연구되었다. 기존 기법이 전체 데이터에 대한 스카이라인을 완전히 구하고, 그 결과를 출력하는데 반하여, 최근접 기반의 스카이라인 질의는 우선적으로 구한 스카이라인부터 바로 출력한다.
- <33> 이와 같은 특성 때문에 Kossman et al.은 자신의 알고리즘을 온라인 알고리즘이라고 정의한다.
- <34> Kossman의 연구는 이전 기법들에 비해 좋은 성능을 보여주지만, 차원이 높아짐에 따라 검색해야 할 영역의 중복이 발생하는 문제가 있다.
- <35> 이러한 문제를 해결하기 위하여 Dimistris et al.은 브랜치/바운드 스카이라인(Branch and Bound Skyline:BBS) 기법을 제안하였다.
- <36> BBS는 R-Tree와 최소 경계 사각형으로 유지되는 데이터 튜플들을 원점으로부터 가까운 순서대로 방문하여 최근접 튜플들을 탐색한다.
- <37> 여기서, 탐색한 튜플은 스카이라인 리스트에 추가하며, 이후 R-Tree의 노드들을 방문할 때, 스카이라인 리스트에 있는 튜플들과 비교하여 지배되는 노드들은 방문하지 않는 가지치기 기법을 사용함으로써 I/O 비용을 매우 감소 시킨다.
- <38> 상기의 기법들은 모두 정적인 데이터를 바탕으로 하나의 질의에 대한 한번의 결과 데이터의 출력만을 고려하였다.
- <39> 그러나, 최근 데이터베이스 분야에서 초점을 맞추고 있는 데이터가 지속적으로 변하는 동적인 환경에서는 결과를 얻기 위하여 매번 새로운 질의 처리를 해야 하기 때문에 현재 지속적으로 변하는 데이터에 최적화된 질의 처리 기법이 필요하게 된다.
- <40> 이에 대하여 Tao et al.은 슬라이딩 윈도우와 R-Tree를 기반으로 스트림 데이터 환경에서의 연속적인 스카이라인 질의를 제안하였으나, 이는 R-Tree를 색인 구조로 사용하였기 때문에 빈번한 데이터 추가, 삭제 및 데이터 값 변경 등으로 색인 갱신에 대한 과비용 문제를 발생시킨다.
- <41> 또한, 매번 튜플의 추가시에 해당 튜플의 스카이라인 여부를 확인하기 위하여 범위 질의 수행에 따른 과비용 문제를 가지고 있다.
- <42> 이의 문제를 해결하기 위하여 Li Tian은 격자 색인을 이용한 연속적인 질의 처리를 제안하였으나, 이 역시 영향 영역을 연결 리스트로 운영함으로써 발생하는 검색 비용의 문제, 원시적인 스카이라인 확인 알고리즘의 비용 문제, 큐를 사용한 셀 방문 알고리즘으로 인한 큐의 비대화 및 중복 회피 비용 문제등이 있다.
- <43> 이와 같이, 종래의 동적인 스카이라인 질의는 R-Tree를 색인으로 사용시 동적인 환경에서 색인 갱신에 대한 과비용 문제가 발생하며, 매번 지배된 영역을 구분하기 위하여 원시적인 알고리즘을 사용하기 때문에 연속적인 질의처리시 질의가 발생할 때마다 상당한 비용이 소모되는 문제점이 있고, 격자 색인을 이용한 연속적인 질의

처리는 원시적인 스카이라인 확인 알고리즘의 과비용 문제 및 큐를 이용한 셀 방문에 따른 큐의 비대화가 야기 되는 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- <44> 따라서, 본 발명이 해결하고자 하는 첫 번째 과제는 튜플들의 속성이 빈번하게 바뀌는 동적인 환경에서도 스카이라인 결과값을 항상 최신의 값으로 갱신할 수 있고, 스카이라인 결과값 추출에 대한 연산량을 줄임으로써 데이터베이스 처리에 있어 연산 비용을 감소시킬 수 있는 셀 상태 값을 이용한 연속 스카이라인 질의 방법을 제공하는 것이다.
- <45> 그리고, 본 발명이 해결하고자 하는 두 번째 과제는 상기 셀 상태 값을 이용한 연속 스카이라인 질의 방법을 적용한 셀 상태 값을 이용한 연속 스카이라인 질의 시스템을 제공하는 것이다.
- <46> 그리고, 본 발명이 해결하고자 하는 세 번째 과제는 상기 셀 상태 값을 이용한 연속 스카이라인 질의 방법을 컴퓨터에서 수행할 수 있도록 프로그램으로 기록된 기록매체를 제공하는 것이다.

과제 해결수단

- <47> 상기 첫 번째 과제를 해결하기 위하여 본 발명은,
- <48> 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정하는 단계; 상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정하는 단계; 상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출하는 단계; 상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정하는 단계; 및 상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신하는 단계를 포함하는 셀 상태값을 이용한 연속 스카이라인 질의 방법을 제공한다.
- <49> 한편, 상기 단위 셀의 대표값은 원점과 가장 가까운 거리에 대응되는 셀의 좌표이며, 상기 단위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현되는 것을 특징으로 한다.
- <50> 그리고, 상기 지역 스카이라인을 선정하는 단계는 상기 단위 셀 내에 포함되는 튜플 들의 다중 속성값을 비교하는 단계; 및 상기 단위 셀 내에 포함되는 튜플들 중에서 하나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정하는 단계를 포함할 수 있다.
- <51> 아울러, 상기 전역 스카이라인을 선정하는 단계는 상기 선정된 지역 스카이라인의 다중 속성값을 비교하는 단계; 및 상기 지역 스카이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정하는 단계를 포함하는 것을 특징으로 한다.
- <52> 그리고, 상기 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역이고, 상기 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 지배 영역을 감싸는 셀들로 이루어진 영역이고, 상기 반 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 작은 대표값을 가지는 셀들로 이루어진 영역이고, 상기 반 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역인 것을 특징으로 한다.
- <53> 한편, 상기 각 셀에 대한 셀 상태값을 설정하는 단계는 상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 SK 상태로 설정하고, 상기 단위 셀이 상기 지배 영역에 해당하면 DOM 상태로 설정하며, 상기 단위 셀이 상기 SK 및 DOM 의 어느 상태에도 속하지 않으면 $FREE$ 상태로 설정하는 단계인 것을 특징으로 한다.

다.

- <54> 그리고, 상기 전역 스카이라인을 갱신하는 단계는 상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하는 단계; 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하는 단계; 및 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정하는 단계를 포함할 수 있다.
- <55> 아울러, 상기 전역 스카이라인을 갱신하는 단계는 상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하는 단계; 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 띠 및 지배 띠를 방문하여 전역 스카이라인의 여부를 결정하는 단계; 및 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플만의 지배 영역 및 지배 띠를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정하는 단계를 포함할 수 있다.
- <56> 한편, 상기 단위 셀 C_{ij} 은 상기 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 상기 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 을 저장하는 것을 특징으로 한다.
- <57> 아울러, 상기 전역 스카이라인을 갱신하는 단계는 상기 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인의 전역 스카이라인의 튜플을 사용자에게 의해 설정된 주기마다 출력하는 단계를 더 포함하는 것을 특징으로 한다.
- <58> 상기 두 번째 과제를 해결하기 위하여 본 발명은,
- <59> 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정하는 셀 대표값 선정부; 상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정하는 지역 스카이라인 선정부; 상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출하는 전역 스카이라인 추출부; 상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정하는 셀 상태값 설정부; 및 상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신하는 스카이라인 갱신부를 포함한다.
- <60> 여기서, 상기 단위 셀의 대표값은 원점과 가장 가까운 거리에 대응되는 셀의 좌표이며, 상기 단위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현되는 것을 특징으로 한다.
- <61> 그리고, 상기 지역 스카이라인 선정부는 상기 단위 셀 내에 포함되는 튜플들의 다중 속성값을 비교하고, 상기 단위 셀 내에 포함되는 튜플들 중에서 하나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정하는 것을 특징으로 한다.
- <62> 한편, 상기 전역 스카이라인 추출부는 상기 선정된 지역 스카이라인의 다중 속성값을 비교하고, 상기 지역 스카

이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정하는 것을 특징으로 한다.

<63> 아울러, 상기 셀 상태값 설정부는 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역을 지배 영역으로 설정하고, 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 지배 영역을 감싸는 셀들로 이루어진 영역을 지배 띠로 설정하며, 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 작은 대표값을 가지는 셀들로 이루어진 영역을 반 지배 영역으로 설정하며, 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역을 반 지배 띠로 선정하여 셀 상태 값을 설정하는 것을 특징으로 한다.

<64> 그리고, 상기 셀 상태값 설정부는 상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 *SK* 상태로 설정하고, 상기 단위 셀이 상기 지배 영역에 해당하면 *DOM* 상태로 설정하며, 상기 단위 셀이 상기 *SK* 및 *DOM*의 어느 상태에도 속하지 않으면 *FREE* 상태로 설정하는 것을 특징으로 한다.

<65> 한편, 상기 스카이라인 갱신부는 상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 *DOM*이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하고, 상기 추가되는 튜플이 포함되는 셀의 상태값이 *FREE* 또는 *SK* 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하며, 상기 추가되는 튜플이 포함되는 셀의 상태값이 *FREE* 또는 *SK* 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정하는 것을 특징으로 한다.

<66> 그리고, 상기 스카이라인 갱신부는 상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 *DOM*이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하고, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 *SK* 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 띠 및 지배 띠를 방문하여 전역 스카이라인의 여부를 결정하며, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 *SK* 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플만의 지배 영역 및 지배 띠를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정하는 것을 특징으로 한다.

<67> 한편, 상기 단위 셀 C_{ij} 은 상기 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 상기 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 을 저장하는 것을 특징으로 한다.

<68> 상기 세 번째 과제를 해결하기 위하여 본 발명은,

<69> 상기 셀 상태 값을 이용한 연속 스카이라인 질의 방법을 컴퓨터에서 수행시킬 수 있는 프로그램으로 기록된 기록매체를 제공한다.

효 과

<70> 본 발명에 의하면, 튜플들의 속성이 빈번하게 바뀌는 환경에서도 스카이라인 결과값을 항상 최신의 값으로 유지할 수 있고, 셀 방문에 있어서의 중복 방문횟수를 줄임으로써 시스템의 성능을 향상시킬 수 있으며, 셀 방문에 따른 유지 비용을 감소시킬 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- <71> 본 발명은 격자 셀의 상태값을 활용한 연속적인 스카이라인 질의를 제안한다.
- <72> 본 발명에 의하면, 셀의 상태 값을 셀 자신이 직접 유지하여 해당 셀에 추가된 튜플이 결과에 영향을 주는지를 확인하는 비용을 $O(1)$ 으로 감소시킬 수 있다.
- <73> 또한, 셀의 상태값을 활용하여 질의처리 시에 방문하여야 하는 셀의 수를 감소시킬 수 있다.
- <74> 한편, 각 셀이 지역 스카이라인을 별도로 유지하게 하여 셀들 간의 튜플 비교시 지역 스카이라인들끼리만 비교하게 함으로써 비교 연산의 횟수를 줄일 수 있다.
- <75> 본 발명은, 한정된 데이터 집합이 지속적으로 변화하는 동적인 환경을 가정하고, 각 데이터는 다중 속성을 가진 튜플이고, 고유의 식별자를 가지고 매 단위시간마다 변화된 튜플에 한해 질의에 대한 데이터로 입력된다.
- <76> 또한, 본 발명에 따른 데이터 베이스는 매 단위 시간마다 입력된 데이터에 한하여 스카이라인을 갱신 처리한 결과를 단위시간별로 출력한다.
- <77> 또한, 본 발명은 데이터 베이스의 처리 결과를 계속 유지하며, 변화된 데이터에 한해서만 점진적으로 갱신시켜, 최종적으로 최신 결과에 대한 데이터를 유지한다.
- <78> 본 발명은 튜플들을 색인하고, 셀 영역들의 상태를 유지하기 위하여 격자 셀 색인 구조를 사용한다.
- <79> 이를 위하여 전체 데이터 도메인을 각 축 모두 동일한 δ 의 크기를 가지는 셀들의 d 차원 배열로 구성한다.
- <80> 본 발명에서는 설명의 편의상 $d=2$ 인 경우로 설정하여 상술하기로 하겠다.
- <81> 셀들은 각 축에서의 위치에 따라 셀 좌표를 가지며, 셀 좌표는 $C_{i,j}$ 로 표기하는데, 이는 x 축 i 번째, y 축 j 번째에 위치하는 셀을 의미한다.
- <82> 격자 셀들은 d 차원 배열로 구성되어 있기 때문에 셀 좌표만으로 직접 접근이 가능하다.
- <83> 각 튜플은 자신의 속성 값을 축의 좌표로 하여 자신이 위치한 셀 좌표를 직접 얻을 수 있다.
- <84> 예를 들면, 튜플 r 의 첫 번째 속성 v_1 을 x 값으로, 두 번째 속성 v_2 를 y 값으로 하여 튜플 r 이 위치할 셀 $C_{i,j}$ 의 인덱스 i, j 는 하기의 수학적 식 1에 따라 직접적으로 연산할 수 있다.

수학적 식 1

<85>
$$i = \lfloor \frac{r.v_1}{\delta} \rfloor, j = \lfloor \frac{r.v_2}{\delta} \rfloor$$

- <86> 한편, 셀들의 전체 집합은 G 로 표시하며, 특정 셀을 기준으로 하는 부분 셀 집합을 표기할 때는 $G(C_{lb_x, lb_y}, C_{rt_x, rt_y})$ 로 표기한다.

- <87> 이는 C_{lb_x, lb_y} 를 가장 왼쪽 아래 셀로 하고, C_{rt_x, rt_y} 를 가장 오른쪽 위 셀로 하는 직사각형 모양의 셀 집합을 의미한다.

- <88> 즉, 이를 기반으로 상기 $G(C_{lb_x, lb_y}, C_{rt_x, rt_y})$ 는 하기의 수학적 식 2와 같이 표현할 수 있다.

수학식 2

<89>
$$G(C_{lb_x, lb_y}, C_{rt_x, rt_y}) = \{C_{x,y} | C_{x,y} \in G, lb_x \leq x \leq rt_x, lb_y \leq y \leq rt_y\}$$

<90> 또한 각 축의 가장 마지막에 있는 셀을 $C_{MX,MY}$ 로 표기한다. 따라서, 셀의 전체집합 G 는 $G(C_{0,0}, C_{MX,MY})$ 와 같다.

표 1

기호	설명
G	격자 셀의 집합
$G(C_{lb_x, lb_y}, C_{rt_x, rt_y})$	$\{C_{x,y} C_{x,y} \in G, lb_x \leq x \leq rt_x, lb_y \leq y \leq rt_y\}$
$C_{MX,MY}$	모든 축의 끝에 위치한 셀
S	전역 스카이라인의 집합
$r.v_i$	튜플 r 의 i 번째 속성 값
$C_{i,j}$	x 축 i 번째 y 축 j 번째 격자 셀
$C_{i,j}.ls$	$C_{i,j}$ 의 지역 스카이라인 리스트
$C_{i,j}.tl$	$C_{i,j}$ 의 지역스카이라인 외의 튜플 리스트
$C_{i,j}.DR$	$C_{i,j}$ 의 지배영역(Dominance Region)
$C_{i,j}.DB$	$C_{i,j}$ 의 지배띠(Dominance Band)
$C_{i,j}.ADR$	$C_{i,j}$ 의 반지배영역(Anti-dominance Region)
$C_{i,j}.ADB$	$C_{i,j}$ 의 반지배띠(Anti-dominance Band)

<91> 상기의 표 1은 본 발명에 적용되는 변수들에 대한 정의를 나타낸 것이다. 이하, 본 발명에 대한 구성은 표 1에
 <92> 개시된 변수를 기반으로 상술하기로 한다.

<93> 실제로, 데이터베이스 시스템은 스카이라인의 튜플의 집합인 질의 처리 결과값을 전역 스카이라인(Global Skyline) 집합 S 에 유지하게 된다.

<94> S 는 최종 결과에 반영되는 튜플의 집합이며, 질의 처리시 추가된 스카이라인 튜플은 S 에 유지하게 된다.

<95> 전역 스카이라인 외에 각 셀 별로 지역 스카이라인(Local Skyline)을 유지하는데, 지역 스카이라인은 하나의 셀
 을 범위로 하는 스카이라인으로써 하나의 셀에서 다수의 튜플들을 유지할 때, 셀들 간 유지하는 튜플들의 비교
 횟수를 줄이기 위하여 사용된다.

<96> 지역 스카이라인에 대한 상세한 내용은 하기에 더욱 상세하게 설명하도록 한다.

<97> 본 발명에서는 셀 기반의 지배 영역과 반 지배 영역을 구성하고, 이를 바탕으로 각 셀의 상태 설정에 따라 연속
 스카이라인 질의 처리를 수행한다.

<98> 그리고, 본 발명은 지역 스카이라인의 유지 방법을 기반으로 셀의 상태와 지역 스카이라인을 활용하여 튜플의
 추가 및 삭제시의 질의처리 알고리즘을 구성한다.

<99> 도 2는 본 발명에 적용되는 셀 기반의 지역 영역 및 지배 띠를 도시한 것이다.

<100> 지배 영역 $C_{i,j}.DR$ 은 임의의 튜플 r 이 셀 $C_{i,j}$ 에 놓여 있다면 튜플 r 의 위치와 상관없이 반드시 지배되는
 영역이다.

<101> 이는 도 2에 도시된 바와 같이, $C_{2,3}$ 에 a 가 놓여있는데, 이때 a 가 셀 $C_{2,3}$ 어디에 놓여 있더라도 항
 상 지배하게 되는 영역 $C_{2,3}.DR$ 은 $G(C_{3,4}, C_{MX,MY})$ (210)가 된다.

- <102> 한편, $C_{ij}.DB$ 는 $C_{ij}.DR$ 을 감싸는 셀들로 C_{ij} 에 놓인 튜플이 지배할 수 있는 최대 영역에서 $C_{ij}.DR$ 을 제외한 영역이다.
- <103> 이는 도 2에 도시된 바와 같이, $C_{2,3}.DR$ (210)을 감싸고 있는 영역이 $C_{2,3}.DB$ (220)이 된다.
- <104> 상기 도 2에 따라, 튜플 r 이 C_{ij} 에 놓여 있다면, 이는 반드시 $C_{ij}.DR$ 영역의 튜플들을 지배한다. 그리고, 튜플 r 은 r 의 위치와 $C_{ij}.DB$ 에 놓인 튜플들의 위치에 따라 $C_{ij}.DB$ 에 놓인 튜플들을 지배한다.
- <105> 또한, $C_{ij}.DR$ 은 C_{ij} 에 놓여 있는 어떤 튜플에 의해 반드시 지배되므로, $C_{ij}.DR$ 안에서의 튜플의 변화는 스카이라인의 질의 결과값에 영향을 미치지 않음을 의미한다.
- <106> 도 3은 본 발명에 적용되는 셀 기반의 반 지배 영역 및 반 지배 띠를 도시한 것이다.
- <107> 반지배 영역 $C_{ij}.ADR$ 은 임의의 튜플이 놓였을 때, C_{ij} 에 놓여진 튜플을 반드시 지배하는 영역이다. 이는 다시 말하면, C_{ij} 가 지배되지 않으면, $C_{ij}.ADR$ 에 어떤 튜플로 없다는 것과 동일하므로, 새 튜플 r 이 추가되었을 때, 놓여진 셀 C_{ij} 가 $FREE$ 상태라면, r 이 스카이라인 인지의 여부를 확인하기 위하여 $C_{ij}.ADR$ 의 셀들은 방문하지 않아도 된다.
- <108> 한편, 도 3에 도시된 바와 같이, a 가 놓여있는 $C_{2,3}$ 에 대한 반 지배 영역 $C_{ij}.ADR$ 은 $G(C_{0,0}, C_{1,2})$ (310)가 된다.
- <109> 반 지배 띠 $C_{ij}.ADB$ 는 임의의 튜플이 놓였을 때 C_{ij} 에 놓여진 튜플을 지배할 수 있는 최대 영역에서 $C_{ij}.ADR$ 을 제외한 영역이다.
- <110> 반 지배띠를 설명하는 도 3을 보면 $C_{2,3}.ADB$ (320)는 $C_{2,3}.ADR$ (310)을 감싸고 있는 영역으로, $C_{2,3}.ADB$ (320)에 놓여진 튜플의 위치에 따라 $C_{2,3}$ 의 튜플을 지배하기 때문에, $C_{2,3}$ 에 튜플이 추가될 때 새 튜플이 스카이라인인지의 여부를 확인하기 위하여 반드시 방문하여야 하는 유일한 영역이 된다.
- <111> 또한 전술한 바에 의하면, $C_{ij}.ADB + C_{ij}.ADR$ 은 어떤 튜플이 놓였을 때 C_{ij} 에 놓여진 임의의 튜플을 지배할 수 있는 최대 영역이 된다.
- <112> 따라서, 어떤 튜플 r 이 셀 C_{ij} 에 있을 때, $C_{ij}.ADB + C_{ij}.ADR$ 내에 있지 않은 튜플 r' 는 절대 r 을 지배할 수 없음을 의미하게 된다.
- <113> 이는 아래와 같이 증명할 수 있다.
- <114> $G - (C_{ij}.ADB + C_{ij}.ADR)$ 영역에 있는 튜플 r' 가 셀 C_{ij} 에 놓여있는 r 을 지배한다고 가정한다. 그러면 지배의 정의에 의하면, r' 가 r 보다 반드시 모든 속성은 작거나 같고 최소한 한 속성은 반드시 작아야 한다.

- <115> 이 정의의 대우는 ' r '의 한 속성이 r' 보다 크다면 r' 는 r 을 지배하지 않는다'가 된다.
- <116> 그러나, r' 가 $G - (C_{ij} \cdot ADB + C_{ij} \cdot ADR)$ 에 놓여 있다는 것은 최소한 C_{ij} 보다 한 속성보다 크게 된다는 것을 의미하므로, 결과적으로 지배하지 않으므로 그 가정은 모순이 되어, 어떤 튜플 r 이 셀 C_{ij} 에 있을 때, $C_{ij} \cdot ADB + C_{ij} \cdot ADR$ 내에 있지 않은 튜플 r' 는 절대 r 을 지배할 수 없다.
- <117> 본 발명은 상기 도 2 및 도 3에서 상술한 정의들에 의해 각각의 셀에 대한 상태값을 설정하는데, 이는 하기와 같은 이점이 있다.
- <118> 첫 번째는, 해당 셀이 지배된 영역인지의 여부를 $O(1)$ 비용으로 알게 된다. 두 번째는 특정 스카이라인 튜플이 배타적으로 지배하는 영역을 알게된다.
- <119> 이를 기반으로 각각의 셀은 세 가지의 상태값 중 하나를 가지게 되는데, 상태값들은 $SK, DOM, FREE$ 와 같다.
- <120> SK 상태값은 해당 셀이 전역 스카이라인 튜플을 가지고 있음을 의미한다. 또한 DOM 상태값은 전술한 지배 영역과 중첩되는 셀들이 가지는 상태값이 된다. 즉, 해당 셀이 어떤 튜플에 의해 지배되었음을 의미한다.
- <121> 또한, SK 와 DOM 어느 상태에도 속하지 않는 셀들은 $FREE$ 상태값을 가진다.
- <122> 셀의 상태를 설명하기 위하여 도 4를 보면 굵은선으로 도시된 셀 $C_{2,2}(410)$ 는 SK 상태로써 $C_{2,2}$ 가 유지하는 지역 스카이라인이 곧 전역 스카이라인임을 의미한다.
- <123> 또한, 회색 셀들은 다른 튜플들에게 지배된 셀(Dominated Cell)로써, 전술한 바와 같이 DOM 상태값을 가지며, 각 축 별로 스카이라인 셀들의 셀 좌표보다 하나씩 큰 셀 좌표를 가진 셀들이다.
- <124> 즉, 도 4에서 $C_{2,2}$ 에 놓인 튜플로부터 지배되는 영역은 $G(C_{3,3}, C_{MX,MY})$ 가 된다. 또한 상기 $C_{2,2}$ 및 $G(C_{3,3}, C_{MX,MY})$ 를 제외한 흰색 셀들은 자유 셀인데, 이 셀들은 스카이라인 셀과 지배된 셀을 제외한 나머지 셀들이다.
- <125> 시스템이 유지하는 전역 스카이라인 외에 각 셀 별로 지역 스카이라인(Local Skyline)을 유지한다. 지역 스카이라인은 하나의 셀이 유지하는 튜플들을 다시 셀 안의 스카이라인으로 구분한 것을 의미한다.
- <126> 지역 스카이라인은 셀들이 유지하는 튜플들 간에 비교할 때, 오직 지역 스카이라인들 끼리만 비교하게 함으로써 연산 비용을 줄이기 위해 사용한다.
- <127> 각각의 셀은 지역 스카이라인 유지를 위하여 두개의 연결 리스트(linked List)를 사용한다.
- <128> 하나의 리스트 $C_{ij} \cdot LS$ 는 지역 스카이라인을 위한 것이고, 나머지 하나 $C_{ij} \cdot TL$ 의 리스트는 지역 스카이라인을 제외한 나머지 튜플들을 저장한다.
- <129> 도 5는 지역 스카이라인을 도시하고 있는데, 이는 각 셀들이 유지하는 검은 점들이 각 셀의 지역 스카이라인이며, 흰색 점들은 지역 스카이라인을 제외한 튜플을 나타낸다.
- <130> 또한, 셀에 튜플이 추가되거나 삭제될 때마다 유지하던 지역 스카이라인은 갱신되어야 할 필요가 있다.
- <131> 본 발명은 튜플의 추가 삭제에 대하여 지역 스카이라인을 항상 최신으로 유지하기 위하여 BNL을 메모리 기반으로 변경하여 사용한다.

- <132> 새 튜플 r 이 어떤 셀 C_{ij} 에 추가될 때 지역 스카이라인 리스트 $C_{ij}.ls$ 의 원소들을 r 과 비교하는데, 이때 r 이 $C_{ij}.ls$ 의 어떤 원소에게 지배된다면, 스카이라인이 되지 못하므로 r 을 $C_{ij}.tl$ 에 추가함으로써 알고리즘이 종료된다.
- <133> 반면, r 이 $C_{ij}.ls$ 의 어떤 원소 r' 를 지배한다면, r' 는 $C_{ij}.ls$ 에서 제거되며, r 이 $C_{ij}.ls$ 에 추가된다.
- <134> 또한, 모든 $C_{ij}.ls$ 의 원소들이 r 을 지배하지 못한다면, 결국에는 비교불가이므로, r 을 $C_{ij}.ls$ 에 추가한다.
- <135> 삭제 시에는 $C_{ij}.ls$ 와 $C_{ij}.tl$ 의 원소들에 대한 BNL을 수행하여 지역 스카이라인을 유지할 수 있다.
- <136> 하기에는 상술한 내용을 기반으로 연속 스카이라인의 질의를 설명한다. 본 발명은 데이터 집합이 변화하는 것을 이벤트로 하여 질의 처리를 시도하며, 다음의 세가지 이벤트를 처리한다.
- <137> 첫 번째는 새 데이터 튜플의 추가이고, 두 번째는 임의의 튜플의 제거이며, 세 번째는 기존 튜플의 값 변경인데, 이 중 세 번째 이벤트인 기존 튜플의 값 변경은 기존 공간질의 기법들과 유사하게 추가와 삭제의 조합으로 처리하게 된다.
- <138> 따라서, 본 발명에서는 새 튜플의 추가와 삭제만을 다루게 된다.
- <139> 한편, 새 튜플 r 이 추가되면, 시스템은 r 이 놓일 셀을 구하고, 구한 셀의 상태값을 확인한다.
- <140> 만약 상태가 DOM 이면, 이미 다른 튜플에게 지배되고 있으므로, 추가된 튜플 r 이 결과에 영향을 미치지 못한다.
- <141> 따라서, 추가 연산 없이 단순히 셀에 r 에 추가하고, 지역 스카이라인을 갱신하는 것으로 알고리즘이 종료된다.
- <142> 반면, 셀의 상태값이 $FREE$ 일 경우에는 추가적인 알고리즘이 필요하다.
- <143> 도 6a 내지 도 6d는 a, b, c 가 스카이라인으로 되어 있고, 새 튜플 h 가 $FREE$ 상태값을 가진 $C_{1,2}$ 에 추가될 때의 알고리즘을 순서대로 설명한 것이다.
- <144> 우선 도 6a에서 보는 바와 같이, h 가 놓여지는 튜플 $C_{1,2}$ 의 상태가 $FREE$ 이므로, 새 튜플 h 가 스카이라인인지를 보장받기 위하여 도 6b에 도시된 바와 같이, 오직 $C_{1,2}.ADB$ 에 포함되는 셀의 튜플들과 비교한다.
- <145> 이때, h 가 $C_{1,2}.ADB$ 의 어떤 튜플에게 지배된다면 h 는 스카이라인이 되지 못하므로, 단순히 셀에 추가되고, 지역 스카이라인을 갱신하는 것으로 알고리즘이 종료된다.
- <146> 그러나, $C_{1,2}.ADB$ 의 어떤 튜플에게도 지배되지 않는다면 h 가 스카이라인임이 보장되므로 h 는 전역 스카이라인 집합 S 에 추가되며, $C_{1,2}$ 를 스카이라인 셀로 설정한다.
- <147> 그리고, 도 6c에 도시된 바와 같이 $C_{1,2}.DB$ 와 $C_{1,2}.DR$ 의 튜플 중 h 에게 지배되는 스카이라인 셀을 찾기 위해 셀들을 방문한다.

- <148> 찾은 스카이라인 셀의 지역 스카이라인의 원소들은 S 에서 제거하며, 방문한 셀들 중에서 $C_{1,2}.DR$ 에 포함되는 셀들은 모두 지배된 셀로 상태값을 설정한다.
- <149> 그러면, 마지막으로 도 6d에 도시된 바와 같이, 튜플의 추가 과정이 종료된 후의 셀들의 상태를 보여준다.
- <150> 또한, 새로 추가된 튜플 r 이 SK 상태의 셀 C_{ij} 에 놓여질 수가 있는데 이런 경우는 r 이 $FREE$ 상태를 가진 셀에 놓여진 경우와 동일하게 처리한다.
- <151> 다만, C_{ij} 는 SK 상태로써 이미 $C_{ij}.DR$ 이 설정되어 있으므로, $C_{ij}.DR$ 은 재 방문할 필요는 없다. 따라서, $C_{ij}.DB$ 만 방문하여 변경된 스카이라인에 대해서만 갱신을 수행하도록 한다.
- <152> 시스템은 유지하던 임의의 튜플 r 에 대한 제거 이벤트가 발생하면, 우선적으로 r 이 놓여진 셀을 구한다.
- <153> 만약 이때 r 이 놓여진 셀의 상태값이 DOM 이라면 r 이 제거 되더라도 결과값에 영향을 미치지 않기 때문에 해당 셀의 지역 스카이라인만 갱신하고 알고리즘은 종료된다.
- <154> 그러나, 만약 r 이 놓여진 셀 C_{ij} 의 상태가 SK 이고, r 을 제거하고난 후 $C_{ij}.ls$ 와 $C_{ij}.tl$ 에 하나 이상의 튜플이 남아 있다면 지역 스카이라인을 갱신하여 지역 스카이라인에 변경 사항이 있는지의 여부를 확인한다.
- <155> 변경 사항이 있다면, 변경된 지역 스카이라인에 한하여 $C_{ij}.ADB$ 를 방문하여, 전역 스카이라인 여부를 확인하고, 이에 따라 $C_{ij}.DB$ 를 방문하여 변경 사항을 S 에 반영한다.
- <156> 반면, r 이 놓여진 셀 C_{ij} 의 상태가 SK 이고, r 을 제거하고난 후, $C_{ij}.ls$ 와 $C_{ij}.tl$ 이 모두 비어있다면, C_{ij} 를 $FREE$ 로 설정하고, $C_{ij}.DB$ 와 $C_{ij}.DR$ 영역에서 오직 r 에게만 지배되어 스카이라인이 되지 못한 튜플, 즉 서브 스카이라인을 찾아야 한다.
- <157> 도 7a 내지 도 7d는 서브 스카이라인을 탐색하는 과정을 도시한 것이다.
- <158> 도 7a를 참조하면, a, b, c 가 전역 스카이라인인 상태에서 a 가 제거되는 상황을 나타내고 있다.
- <159> 우선 시스템은 a 를 $C_{2,2}$ 에서 삭제한다. 그리고, $C_{2,2}$ 의 상태값이 SK 이기 때문에 $C_{2,2}.ls$ 와 $C_{2,2}.tl$ 을 확인한다.
- <160> 그러면, $C_{2,2}.ls$ 와 $C_{2,2}.tl$ 모두 비어 있기 때문에 해당 셀에서 서브 스카이라인을 찾을 수 없으므로 a 가 지배하던 영역에서 서브 스카이라인을 찾아야 한다.
- <161> 따라서, $C_{2,2}$ 를 $FREE$ 로 상태값을 변경하고 시스템은 a 가 지배하던 영역의 셀을 방문하여 오직 a 에게만 지배되던 튜플을 탐색한다.
- <162> 그리고, $C_{2,2}$ 에 있던 a 가 지배하는 영역과 겹쳐지는 최대 영역과 중첩되는 $C_{2,2}.DB + C_{2,2}.DR$ 을 방문해야 한다.

- <163> 그러나, $C_{2,2}.DB + C_{2,2}.DR$ 에는 다른 스카이라인이 지배하는 영역이 포함될 수도 있다.
- <164> 오직 a 에 의해 지배되던 튜플들을 찾기 위해서는 a 에게만 지배되는 셀들을 방문하는 것이 더 셀의 방문 횟수를 줄일 수 있다.
- <165> 오직 a 가 지배하던 영역의 셀만을 방문하기 위해서는 a 가 지배된 상태로 설정하지 않았음에도 a 의 서브 스카이라인을 찾기 위해 반드시 방문해야 하는 $C_{2,2}.DB$ 를 활용하여 오직 a 가 지배하는 영역을 골라낼 수 있다.
- <166> 도 7b는 a 가 지배하던 영역 중에서 다른 스카이라인에게 지배되지 않는, 즉 오직 a 에게만 지배되던 영역을 알기 위하여 $C_{2,2}.DB$ 를 방문하는 것을 도시하고 있다.
- <167> 이때, x 축 범위를 알기 위하여 $C_{2,2}$ 부터 시작하여 $C_{6,2}$ 까지 이동하여 x 축의 범위가 $2 \leq x < 6$ 임을 구하고, $C_{2,2}$ 부터 시작하여 $C_{6,2}$ 까지 이동하여 $2 \leq y < 6$ 까지임을 구한다.
- <168> 각 축별 범위를 구할 때, 축 별 마지막 셀을 취하며, 또한 지배된 셀의 이전 셀을 각 축 별 범위로 가지게 되는 이유는 지배된 셀이 이미 다른 튜플에게 지배되고 있으므로 서브 스카이라인이 존재할 수 없기 때문이다.
- <169> 도 7c는 구한 각 축 별의 범위를 구한 뒤, 서브 스카이라인을 구하기 위한 셀 방문을 도시하고 있다.
- <170> 도 7c를 보면 a 가 놓인 셀 $C_{2,2}$ 로부터 우선 셀의 y 좌표를 1씩 증가시켜 y 의 범위의 한계인 $C_{2,5}$ 까지 방문한 후, 다시 시작하여 x 를 1 증가시키고, y 값을 초기화함으로써 $C_{3,2}$ 로부터 $C_{3,5}$ 까지 방문하게 하는 식의 셀 방문을 각 축 모든 범위 한계에 다다를 때까지 반복한다.
- <171> 이때, 서브 스카이라인 후보를 발견하면 발견된 후보를 후보 리스트에 넣고, 발견된 후보가 위치한 셀로부터 새 범위를 구할 수 있다.
- <172> 즉, 도 7c에서 $C_{3,4}$ 를 방문하여 e 를 서브 스카이라인 후보로 추가한 후에는 y 축의 범위를 $2 \leq y \leq 4$ 로 조정한다.
- <173> e 를 발견한 이후에 y 축의 범위를 갱신하는 이유는 e 의 지배 영역을 방문하지 않기 위한 것으로, 우선 e 를 발견한 후에는 방문하는 셀들, 즉 $G(C_{4,2}, C_{MX,MY})$ 의 튜플은 e 를 지배할 수 없다.
- <174> 따라서, e 는 반드시 스카이라인이 되며, e 에 대한 지배영역 $G(C_{4,5}, C_{MX,MY})$ 은 서브 스카이라인이 존재할 수 없으므로, e 의 지배영역은 방문할 필요가 없게 된다.
- <175> 그러면, 도 7d에서 볼 수 있는 바와 같이, 튜플의 제거를 완료한 후의 새로운 셀 상태값이 완성되게 된다.
- <176> 도 8은 본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 방법의 흐름도를 도시한 것이다.
- <177> 상기 도 8의 설명에 있어서 중복된 상술은 과감히 생략하기로 한다.
- <178> 우선, 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정한다(810 과정).
- <179> 상기 단위 셀의 대표값은 원점과 가장 가까운 거리에 대응되는 셀의 좌표인 것을 특징으로 하며, 이는 상기 단

위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현될 수 있다.

- <180> 그 다음, 상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정한다(820 과정).
- <181> 지역 스카이라인을 선정하기 위하여, 상기 단위 셀 내에 포함되는 튜플 들의 다중 속성값을 비교하고, 상기 단위 셀 내에 포함되는 튜플들 중에서 하나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정할 수 있다.
- <182> 한편, 상기 단위 셀은 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 를 저장할 수 있다.
- <183> 그 다음, 상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출한다(830 과정).
- <184> 전역 스카이라인을 선정하기 위하여, 상기 선정된 지역 스카이라인의 다중 속성값을 비교하고, 상기 지역 스카이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정할 수 있다.
- <185> 그 다음, 상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정한다(840 과정).
- <186> 여기서, 상기 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역이고, 상기 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 지배 영역을 감싸는 셀들로 이루어진 영역이고, 상기 반 지배 영역은 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 작은 대표값을 가지는 셀들로 이루어진 영역이고, 상기 반 지배 띠는 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역인 것을 특징으로 한다.
- <187> 한편, 셀 상태값을 설정하기 위하여 상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 SK 상태로 설정하고, 상기 단위 셀이 상기 지배 영역에 해당하면 DOM 상태로 설정하며, 상기 단위 셀이 상기 SK 및 DOM 의 어느 상태에도 속하지 않으면 $FREE$ 상태로 설정할 수 있다.
- <188> 마지막으로, 상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신한다(850 과정).
- <189> 여기서, 상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신하고, 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하게 된다.
- <190> 그리고, 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 띠에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정할 수 있다.
- <191> 한편, 상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신한다.

- <192> 만약, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 띠 및 지배 띠를 방문하여 전역 스카이라인의 여부를 결정한다.
- <193> 그러나, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플만의 지배 영역 및 지배 띠를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정할 수 있다.
- <194> 한편, 상기 전역 스카이라인을 갱신하는 단계(850 과정)는 상기 다중 속성값을 가지는 복수 개의 튜플로 구성된 데이터 도메인의 전역 스카이라인의 튜플을 사용자에게 의해 설정된 주기마다 출력하는 과정을 더 포함할 수 있다.
- <195> 도 9는 본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 시스템의 블록도를 도시한 것이다.
- <196> 도 9를 참조하면, 본 발명에 따른 연속 스카이라인 질의 시스템은 셀 대표값 선정부(910), 지역 스카이라인 선정부(920), 전역 스카이라인 추출부(930), 셀 상태값 설정부(940) 및 스카이라인 갱신부(950)로 이루어질 수 있다.
- <197> 셀 대표값 선정부(910)는 다중 속성값을 가지는 복수 개의 튜플(Tuple)로 구성된 데이터 도메인을 각 축 모두 동일한 크기를 가지는 격자형의 단위 셀들의 다차원 배열로 구성하고, 단위 셀의 대표값을 선정한다.
- <198> 여기서, 상기 단위 셀의 대표값은 원점과 가장 가까운 거리에 대응되는 셀의 좌표이며, 상기 단위 셀이 x 축 i 번째, y 축 j 번째 위치할 경우, C_{ij} 로 표현될 수 있다.
- <199> 지역 스카이라인 선정부(920)는 상기 다차원 배열로 구성된 단위 셀 별로 자신의 셀에 포함된 튜플로부터 각 차원의 축 값의 크기를 비교하여 지역 스카이라인을 선정한다.
- <200> 상기 지역 스카이라인 선정부(920)는 상기 단위 셀 내에 포함되는 튜플들의 다중 속성값을 비교하고, 상기 단위 셀 내에 포함되는 튜플들 중에서 하나 이상의 가장 작은 속성값을 가지는 튜플을 지역 스카이라인으로 선정한다.
- <201> 전역 스카이라인 추출부(930)는 상기 선정된 지역 스카이라인을 기반으로, 상기 지역 스카이라인이 속한 셀의 대표값의 크기를 비교하여 상기 데이터 도메인의 전역 스카이라인을 추출한다.
- <202> 상기 전역 스카이라인 추출부(930)는 상기 선정된 지역 스카이라인의 다중 속성값을 비교하고, 상기 지역 스카이라인 중에서 적어도 하나 이상의 가장 작은 속성값을 가지는 튜플을 전역 스카이라인으로 선정한다.
- <203> 셀 상태값 설정부(940)는 상기 전역 스카이라인에 따라 전역 스카이라인 각각의 지배 영역, 지배 띠, 반 지배 영역 및 반 지배 띠를 구성하여 각 셀에 대한 셀 상태값을 설정한다.
- <204> 상기 셀 상태값 설정부는 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 큰 대표값을 가지는 셀들로 이루어진 영역을 지배 영역으로 설정한다.
- <205> 그리고, 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며 상기 지배 영역을 감싸는 셀들로 이루어진 영역을 지배 띠로 설정한다.
- <206> 한편, 상기 추출된 전역 스카이라인의 셀의 대표값에 대응되는 각 축의 값보다 모든 축의 값이 작은 대표값을 가지는 셀들로 이루어진 영역을 반 지배 영역으로 설정한다.
- <207> 그리고, 상기 추출된 전역 스카이라인이 포함되는 셀을 포함하며, 상기 반 지배 영역을 감싸는 셀들로 이루어진 영역을 반 지배 띠로 설정함으로써 셀 상태값을 결정하게 된다.
- <208> 한편, 상기 셀 상태값 설정부(940)는 상기 단위 셀이 상기 전역 스카이라인의 튜플을 가지고 있으면 SK 상태로 설정하고, 상기 단위 셀이 상기 지배 영역에 해당하면 DOM 상태로 설정하며, 상기 단위 셀이 상기 SK

및 DOM 의 어느 상태에도 속하지 않으면 $FREE$ 상태로 설정할 수 있다.

- <209> 한편, 본 발명에 적용되는 단위 셀 C_{ij} 은 상기 지역 스카이라인을 저장하는 지역 스카이라인 리스트 $C_{ij}.ls$ 및 상기 지역 스카이라인을 제외한 나머지 튜플을 저장하는 셀 튜플 리스트 $C_{ij}.tl$ 을 저장할 수 있다.
- <210> 스카이라인 갱신부(950)는 상기 셀 상태값이 설정된 데이터 도메인에 새로운 튜플이 추가 또는 삭제되면, 상기 추가 또는 삭제된 튜플이 포함되는 셀의 셀 상태값 및 지역 스카이라인에 따라 상기 전역 스카이라인을 갱신한다.
- <211> 상기 스카이라인 갱신부(950)는 상기 데이터 도메인에 새로운 튜플이 추가될 경우, 상기 추가되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신한다.
- <212> 그리고, 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 때에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하는 경우, 상기 추가되는 튜플이 포함되는 셀의 지역 스카이라인 정보만 갱신하게 된다.
- <213> 또한, 상기 추가되는 튜플이 포함되는 셀의 상태값이 $FREE$ 또는 SK 이고, 상기 추가되는 튜플이 존재하는 셀의 반 지배 때에 존재하는 튜플 중에서 각 축의 값보다 모든 축의 값의 작은 튜플이 존재하지 않는 경우, 상기 추가되는 튜플을 스카이라인 정보로 갱신하고, 상기 추가되는 튜플이 존재하는 셀을 스카이라인 셀로 설정할 수 있다.
- <214> 한편, 상기 스카이라인 갱신부(950)는 상기 데이터 도메인에 특정 튜플이 삭제될 경우, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 DOM 이면 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보만 갱신한다.
- <215> 그리고, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하는 경우, 상기 삭제되는 튜플이 포함되는 셀의 지역 스카이라인의 정보를 갱신하고, 상기 지역 스카이라인의 갱신된 정보에 따라 상기 셀의 반 지배 때 및 지배 때를 방문하여 전역 스카이라인의 여부를 결정하게 된다.
- <216> 또한, 상기 삭제되는 튜플이 포함되는 셀의 상태값이 SK 이고, 상기 삭제되는 튜플이 포함되는 셀에 다른 튜플이 존재하지 않는 경우, 상기 삭제되는 튜플만의 지배 영역 및 지배 때를 방문하여 새로운 스카이라인을 탐색하고, 상기 스카이라인을 기준으로 상기 데이터 도메인의 셀 상태값을 재설정하게 된다.
- <217> 본 발명의 성능 평가를 위하여 같은 환경에서 제안된 Li Tian et al.의 CMSQ와 비교 성능 평가를 수행하였다.
- <218> 시뮬레이션은 최초 초기화를 끝낸 후 속성 값 변화에 따른 길의 처리 비용을 측정하며, 튜플 수의 증가와 데이터 증가에 따른 연산 비용(CPU cost), 독립(Independent) 데이터 분포를 바탕으로 격자 입도의 증가에 따른 연산 비용을 평가 하였다.
- <219> 시뮬레이션 매개변수는 하기의 표 2와 같으며 굵은 색 글씨의 매개 변수는 기본 매개 변수로 해당 매개 변수에 대한 시뮬레이션이 아닌 경우, 사용된 기본 매개변수 값을 의미한다.

표 2

매개변수	값
격자 입도 (granularity)	1, 25, 50, 100, 125, 250, 500, 1000
튜플의 수	1k, 5k, 10k, 15k, 20k, 25k, 30k, 35k, 45k, 50k, 55k, 60k, 65k, 70k, 75k, 80k, 85k, 90k, 100k
차원	2
데이터 분포	Independent, Correlated, Anti-correlated
속성 값 변경 비율	10%
속성 값 변경 범위	1 - 50
격자 범위	10000 × 10000

<220>

<221> 도 10a 내지 도 10c는 튜플의 수에 따른 연산 시간을 비교 측정한 것이다.

<222> 여기서 튜플 수를 1000개에서 100000개까지 5000개 단위로 변화시키면서 연속적으로 질의하여 얻은 평균 처리 시간을 그래프로 표현하였다.

<223> 도 10a와 도 10b를 보면, 독립(Independent)과 연관(Correlated) 데이터 분포에서 전체적으로 본 발명에 있어서의 제안 기법의 처리 시간이 더 짧은 것을 알 수 있다.

<224> CMSQ는 튜플이 증가함에 따라 처리시간이 크게 증가하는 것을 볼 수 있는데, 그 이유는 영향 지역을 리스트로 유지하므로 매 튜플 처리시마다 매번 영향 지역을 확인하기 위해 리스트를 순회하므로 큰 비용이 드는 것과 추가된 튜플의 스카이라인 확인 여부를 시스템이 유지하는 모든 스카이라인과 비교하는 것 때문이라고 분석되어 진다.

<225> 도 10c는 반 상관(Anti-Correlated) 데이터 분포를 기반으로 연산 비용을 시뮬레이션을 수행한 결과이다.

<226> CMSQ의 경우, 추가된 튜플이 스카이라인 인지의 여부를 확인하기 위하여 시스템이 유지하는 모든 스카이라인과 비교하기 때문에 스카이라인의 갯수가 성능에 아주 중요한 요소가 된다.

<227> 그런데, 반 상관 분포는 전체 튜플 수에 비해 스카이라인의 비율이 많은 특징을 가지고 있다.

<228> 또한 튜플이 증가함에 따라 스카이라인 수도 많이 증가하게 되고 결과적으로 본 발명과 비교하여 기존 튜플의 삭제 시는 지배 때 방문을 통하여 방문할 셀의 범위를 특정 셀의 주변으로 범위를 제한하므로 연산 비용을 감소 시킨다.

<229> 또한, 본 발명에 의하면 지역 스카이라인 유지를 통하여 튜플이 많더라도 지역 스카이라인 튜플끼리만 비교하므로, 비교 횟수가 줄게되어 증가하는 튜플 수에 대한 연산 비용의 증가가 감소하게 된다.

<230> 실제로, 격자 셀 색인의 특성상 튜플의 수가 많아지면, 한 셀이 유지해야 하는 튜플의 수가 많아짐으로 인한 성능저하가 발생한다.

<231> 그러나, 격자의 입도를 늘려 셀에 가중되는 부하를 분산할 수 있다.

<232> 그러나, 알고리즘의 특성상 격자의 입도 자체가 알고리즘의 성능에 영향을 줄 수도 있기 때문에 여기서는 격자의 입도에 따른 알고리즘의 성능을 비교 분석한다.

<233> 도 11은 격자의 입도(granularity) 변화에 따른 처리 시간을 CMSQ와 비교 측정한 그래프이다.

<234> 두 기법 모두 격자 셀을 d 차원 배열로 구현하여 튜플의 색인 시간은 크게 다르지 않으나, CMSQ는 영향 지역을 리스트로 유지하기 때문에 입도가 증가함에 따라 영향 지역을 유지하는 리스트의 길이도 길어진다.

<235> 따라서, 입도가 증가함에 따라 급격하게 처리 시간이 증가하는 것을 볼 수 있다.

<236> 그러나, 본 발명에 의하면, 격자 입도가 큰 경우에도 셀의 지배 여부를 셀 자체가 유지하기 때문에 입도에 대해 일정한 처리 시간 증가를 보여준다.

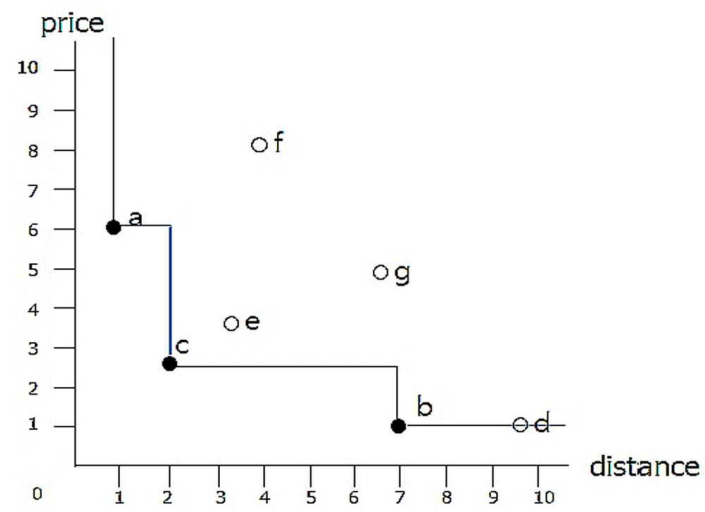
- <237> 또한, 입도가 낮은 환경에서도 CMSQ보다 효율적인 것을 볼 수 있는데 이는 지역 스카이라인이 하나의 셀에 많은 튜플이 유지되더라도 결국 비교하여야 할 튜플의 수를 제한되게 하므로, 셀들이 유지하는 튜플들을 비교할 때 처리 시간을 감소시키는 것을 알 수 있다.
- <238> 본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 방법은 소프트웨어를 통해 실행될 수 있다. 소프트웨어로 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.
- <239> 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- <240> 본 발명은 도면에 도시된 일 실시예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며, 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시예의 변형이 가능하다는 점을 이해할 것이다.
- <241> 그러나, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다.
- <242> 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

도면의 간단한 설명

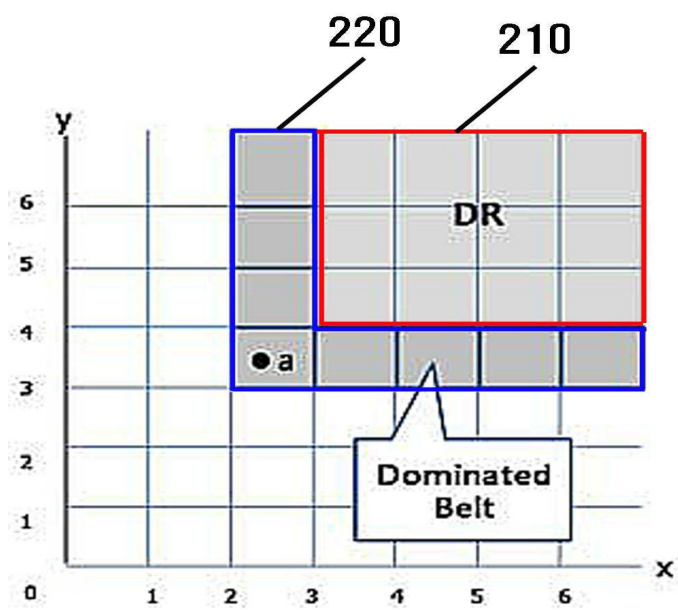
- <243> 도 1은 스카이라인 질의에 적용되는 속성값 비교 그래프를 도시한 것이다.
- <244> 도 2는 본 발명에 적용되는 셀 기반의 지역 영역 및 지배 띠를 도시한 것이다.
- <245> 도 3은 본 발명에 적용되는 셀 기반의 반 지배 영역 및 반 지배 띠를 도시한 것이다.
- <246> 도 4는 본 발명에 적용되는 스카이라인 셀과 지배된 셀을 도시한 것이다.
- <247> 도 5는 본 발명에 적용되는 전역 및 지역 스카이라인을 도시한 것이다.
- <248> 도 6a 내지 도 6d는 본 발명의 일 실시예에 따른 새 튜플이 추가될 때의 알고리즘을 도시한 것이다.
- <249> 도 7a 내지 도 7d는 본 발명의 일 실시예에 따른 서브 스카이라인을 탐색하는 과정을 도시한 것이다.
- <250> 도 8은 본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 방법의 흐름도를 도시한 것이다.
- <251> 도 9는 본 발명에 따른 셀 상태 값을 이용한 연속 스카이라인 질의 시스템의 블록도를 도시한 것이다.
- <252> 도 10a 내지 도 10c는 튜플의 수에 따른 연산 시간을 비교 측정한 것이다.
- <253> 도 11은 격자의 입도(granularity) 변화에 따른 처리 시간을 CMSQ와 비교 측정한 그래프이다.

도면

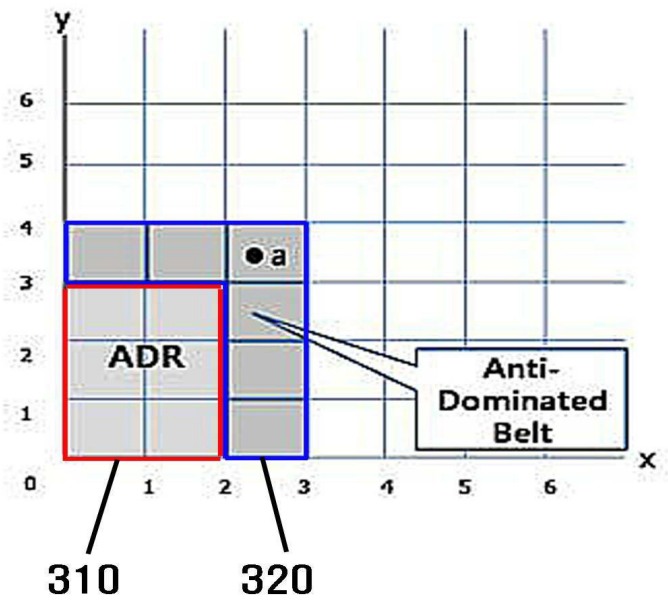
도면1



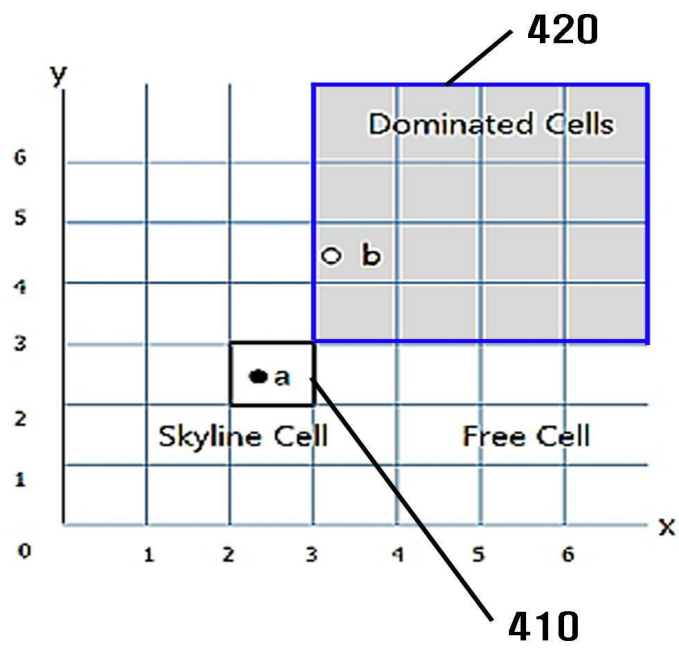
도면2



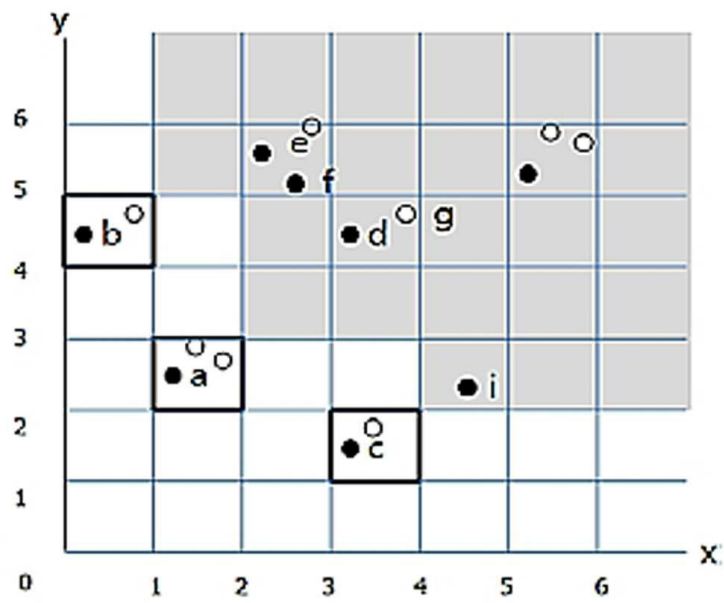
도면3



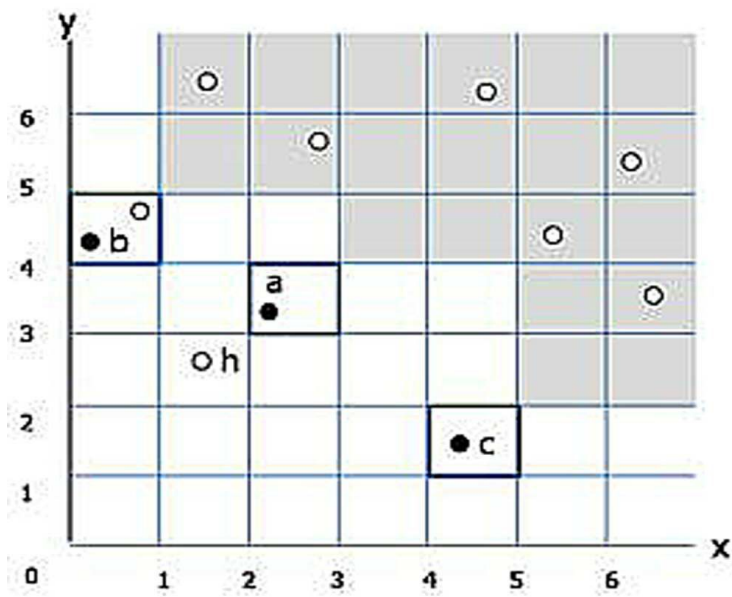
도면4



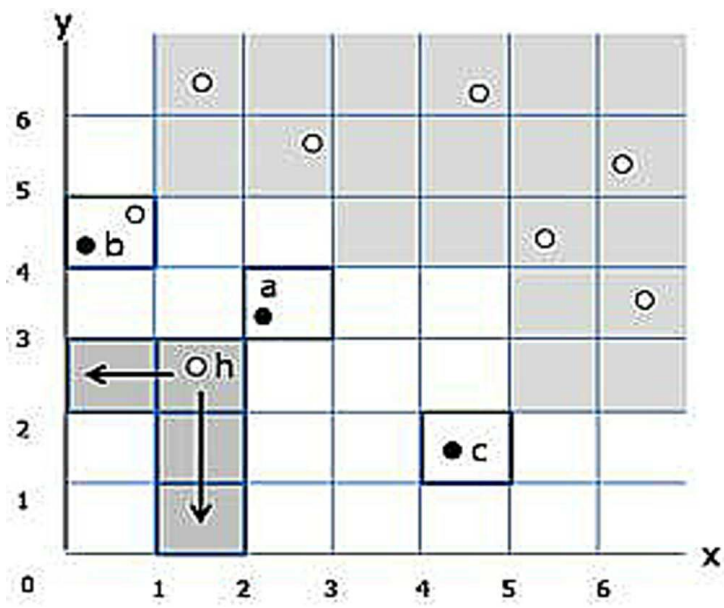
도면5



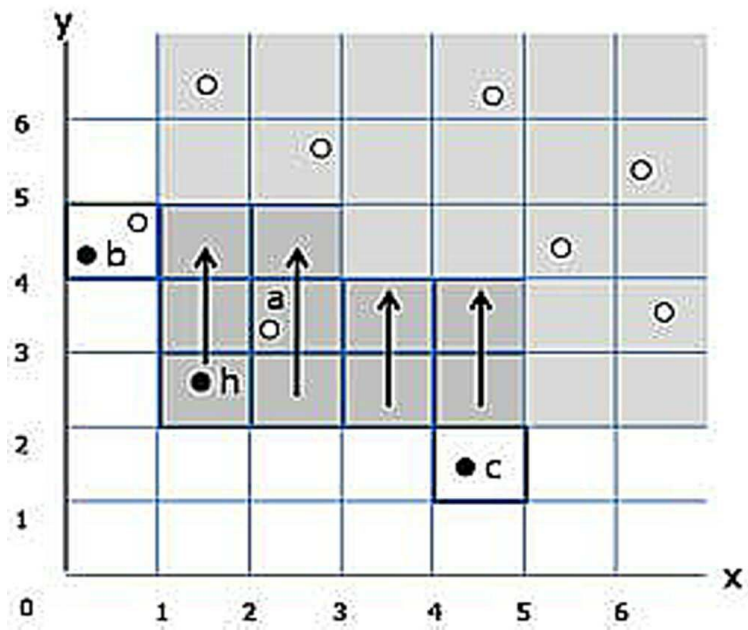
도면6a



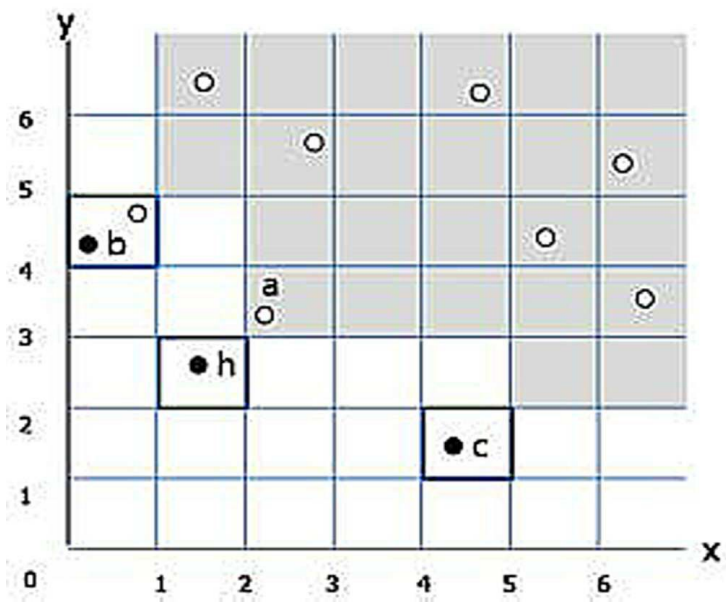
도면6b



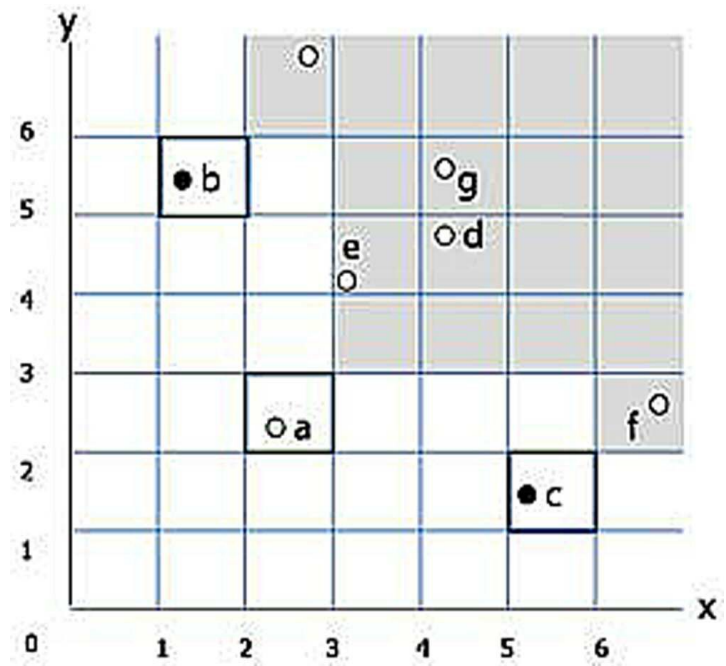
도면6c



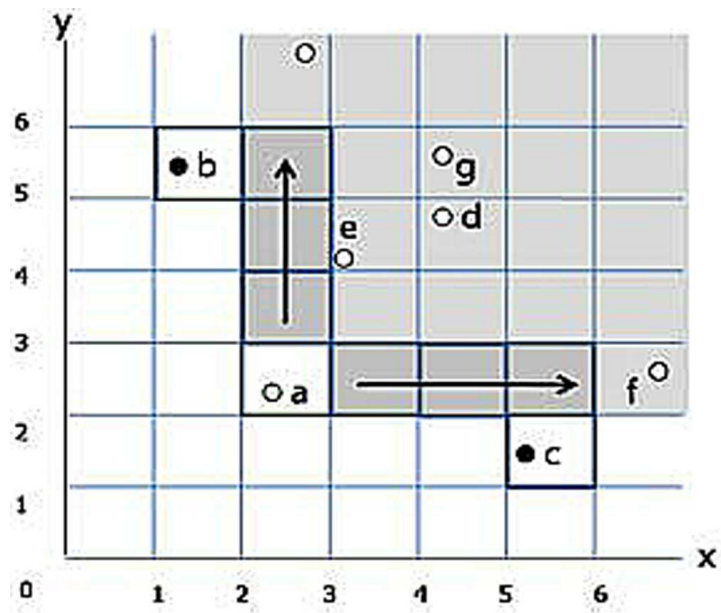
도면6d



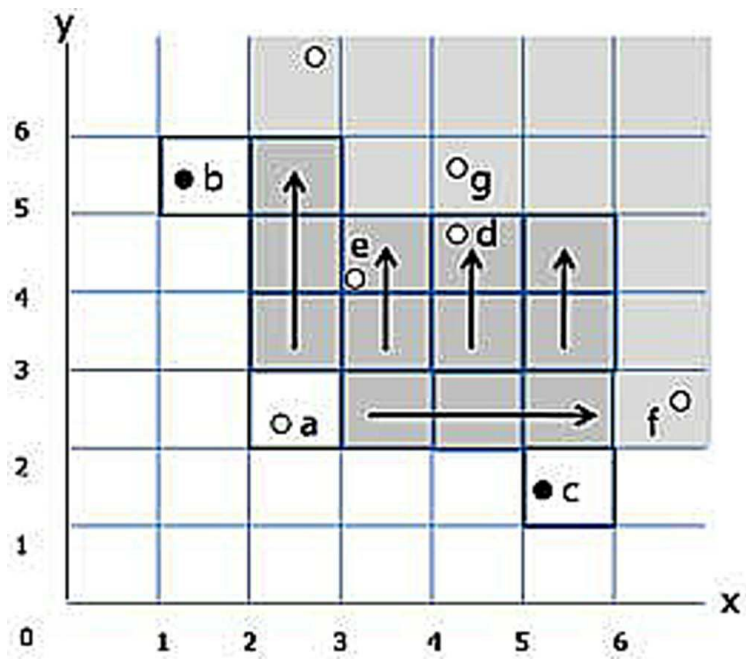
도면7a



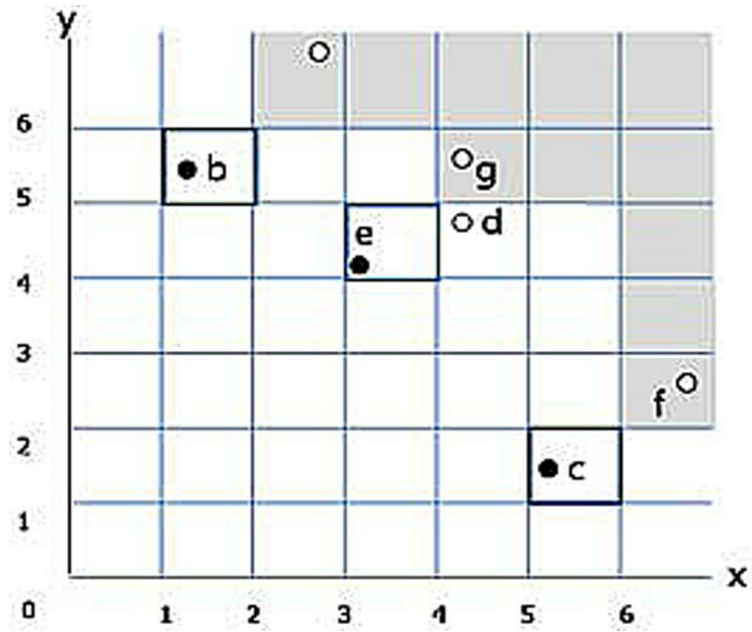
도면7b



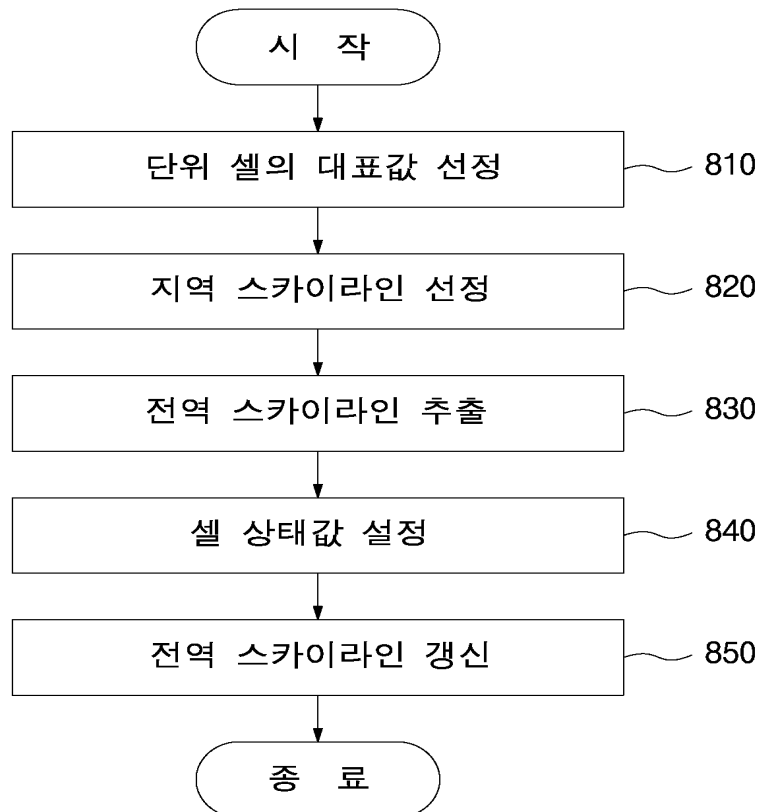
도면7c



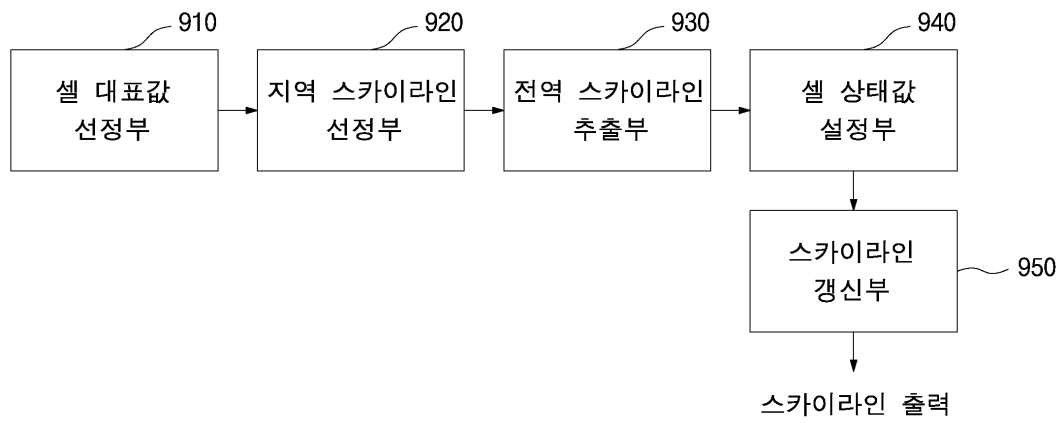
도면7d



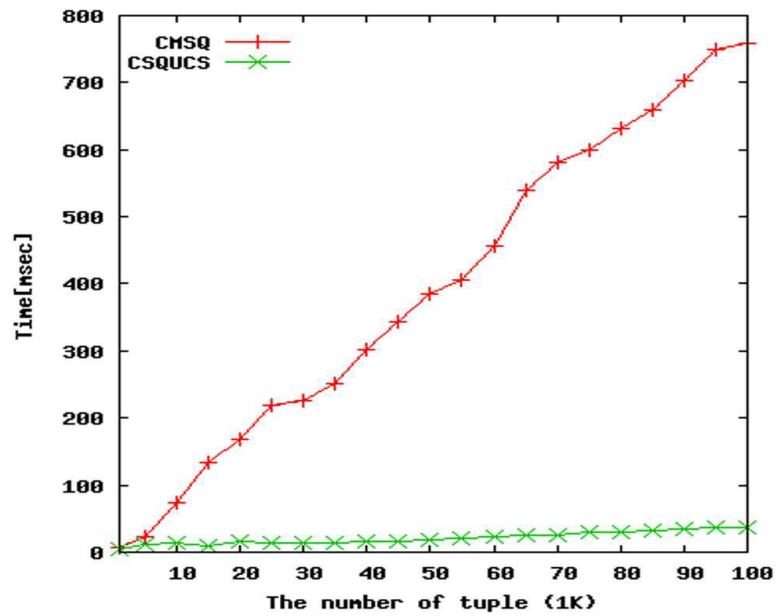
도면8



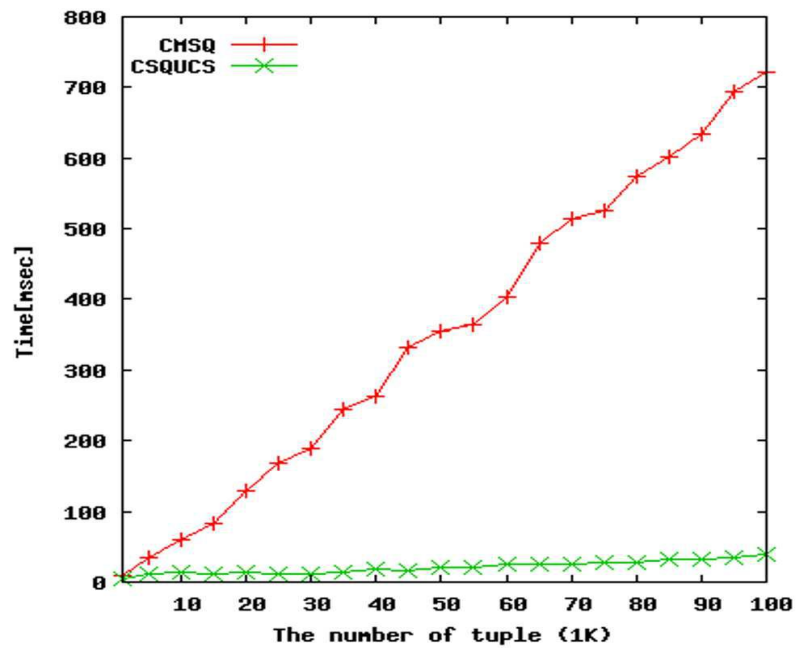
도면9



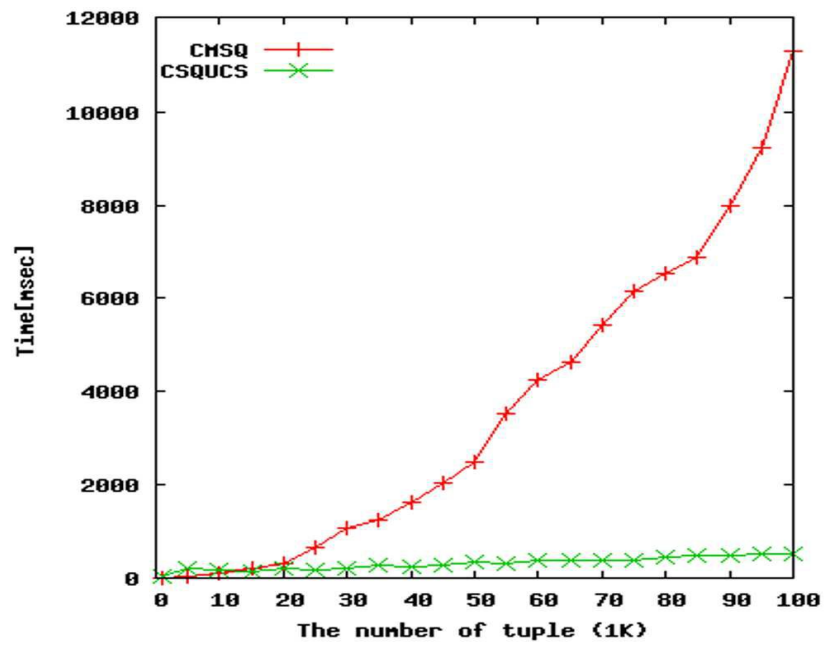
도면10a



도면10b



도면10c



도면11

