

# R<sup>+</sup>-tree를 활용한 지하 단층 분석 시스템 개발 연구

방 갑 산  
한성대학교 정보시스템공학과

## Sub-surface Fault Analysis System Development Using R<sup>+</sup>-tree

Kapsan Bang  
Dept of Information System Engineering, Hansung University

### 요 약

최근의 비약적인 하드웨어와 소프트웨어의 발달에 따라 효율적인 대량의 정보처리가 가능해졌고 특히 다차원의 공간정보처리에 대한 많은 연구가 진행중이다. 공간자료 구조를 활용한 지리정보시스템(GIS)에 의한 공간정보의 분석과 처리는 비약적인 발전을 하고 있다. 그러나 대부분의 지리정보시스템의 응용분야는 지상에 존재하는 것을 대상으로 하고 있다. 지리정보시스템을 활용한 지하정보에 대한 연구는 지하정보의 비가시적인 특성상 아직은 초기단계라고 할 수 있다. 특히 지하 지층 분포를 체계적으로 분석하기 위한 연구는 지하 광물자원에 대한 정보 또는 지상의 구조물에 대한 안전성 평가에 대단히 중요한 역할을 한다. 본 논문은 지하 지층을 공간자료구조를 활용하여 인식하는 시스템 개발 방법을 제안하고 이와 관련된 사항을 기술하였다. 공간자료 구조를 활용하여 지하지층을 인식하는 시스템의 개발은 지질 전문가의 해석을 위한 전처리 시스템으로서 중요한 역할을 할 것이다.

### 1. 서론

최근의 컴퓨터 하드웨어와 소프트웨어의 비약적인 발전에 의해 대용량의 다차원 공간정보의 효율적인 처리가 가능해짐에 따라 공간데이터분야가 최근에 급속히 발전되어가고 있다. 이러한 활발한 연구의 결과로 지형공간정보의 효율적인 처리가 가능한 지리정보시스템(GIS)과 같은 분야가 비약적인 발전을 하고 있다. 대부분의 지리정보시스템의 응용분야는 지상에 존재하는 대상물을 처리하는 것을 목적으로 개발되고 있다. 반면, 지하지형에 대한 체계적인 분석과 연구 그리고 축적된 정보를 효율적으로 저장/활용하는 시스템에 대한 연구는 지하정보의 비가시적인 특성상 아직은 초기단계라 할 수 있다. 지하 광물자원을 탐사하기 위해 일부 지역에 제한된 수의 시추공을 굴착하거나 경도를 통해 직접 지하 광물과 일부 구조를 파악 할 수도 있으나 이는 매우 제한적인 범위에서

행하여지므로 전체의 광물에 대한 분포나 특히 지하 지형의 전반적인 구조를 파악하는데는 한계가 있다. 따라서 보다 넓은 지역의 지하 구조를 파악하기 위해 탄성파가 많이 활용되고 있다. 탄성파를 통해 얻어진 자료를 처리하여 다양한 분야에 필요한 많은 정보를 얻을 수 있으나 처리과정에서 지질전문가의 해석이 요구된다. 또한 얻어지는 정보의 양이 방대하고 3차원의 정보로 확대하여 해석할 경우 전문가의 주관적인 해석에 의한 데이터 처리는 한계를 가질 수밖에 없다. 지하구조 가운데서 지층의 단층상태를 시스템으로 인식하는 작업은 거의 시도되지 않고 있는 상태이다. 본 논문은 지하 지층을 공간자료구조를 활용하여 인식하는 시스템 개발과 관련된 기술적인 방법을 제안하고 이와 관련된 사항을 기술하였다.

## 2. 관련연구

### 2.1 탄성파 데이터

탄성파는 암석이나 광물의 탄성의 차에 의한 전파 속도를 측정하고, 이를 해석하는 탐사법이며, 탄성파 탐사법이 가장 광범위하게 이용되는 분야는 석유탐사 분야, 지반 및 복잡한 지질구조를 조사할 때도 사용한다. 탄성파 탐사에는 반사법 탄성파(reflection survey)와 굴절법 탄성파(refraction survey) 탐사의 두 가지가 있다.

본 논문에서 사용된 연구지역은 미국 north-central Texas의 Fort Worth Basin에 위치한 Boonsville 천연가스 field이다 (그림 1).

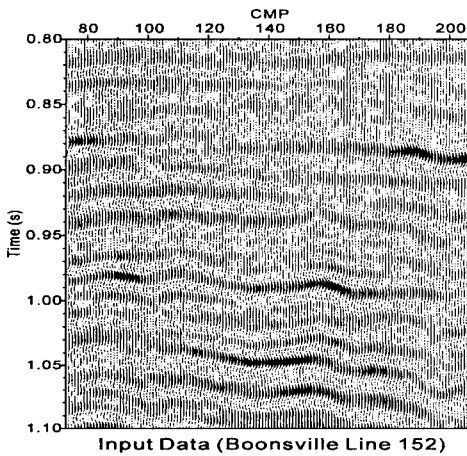


그림 1. 탄성파 데이터

### 2.2 공간자료구조

공간자료구조[5, 6]의 연구가 처음으로 시작된 것은 70년대 후반이지만 본격적인 연구는 80년대 후반부터 시작되어 현재 이 분야에 대한 연구가 활발하게 진행이 되고있다. 그러나, 1차원 자료구조 분야와는 달리 공간자료구조 분야에서는 어느 구조도 모든 응용분야에서 안정된 성능을 보여주고 있지 못하고 있다. 그 이유는 응용분야에 따라 데이터의 분포와 질의 형태가 다르며 공간자료구조에서 공간분할에 고려되어야 할 요소가 많기 때문이다. 예를 들어 GIS와 VLSI에서의 데이터의 분포나 질의 특성의 차는 너무나도 다르다. GIS 데이터의 처리에 있어 커다란 복잡한 객체는 작은 객체들로 분해가 되고 각 객체들은 다른 객체들에 대한 참조 정보를 가지고 있어 이것으로 공간정보를 형성한다. 그러나 VLSI 데이터는 경계를 공유하는 데이터의 패턴이 없기 때문에 이런

참조정보가 없다. 공간데이터를 활용하는 시스템을 구축하는데 있어서 공간자료구조의 선택이 전체시스템의 성능에 결정적인 영향을 미친다. 1차원 데이터를 사용하는 system에서 1차원 자료구조의 선택처럼 전체의 성능에 별로 큰 영향을 주지 않는 것과 달리, 응용분야에 따라 데이터의 성격과 데이터에 대한 질의 형태가 특정한 어느 공간자료구조를 더 선호하도록 할 수 있기 때문이다.

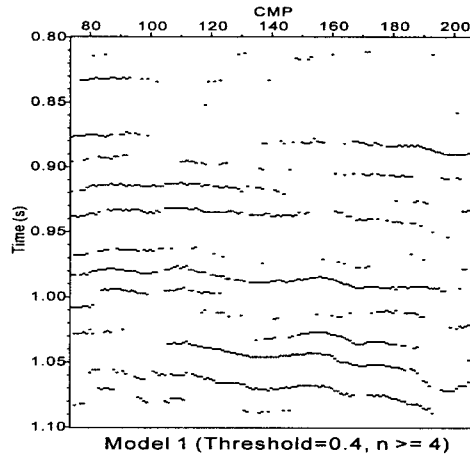


그림 2. 그림 1(threshold= 0.4)에 대한 thinning

이 논문에서 제안하는 지하단층 인식 시스템에 가장 적절한 공간자료구조는 공간상에서 포인트 데이터를 구성하고 검색하는데 가장 효과적인 자료구조로서 알려져 있는  $R^+$ -tree[5, 6]이다.  $R^+$ -tree는  $R$ -tree[2]의 변형이다.  $R^+$ -tree의 노드분할 알고리즘은 노드분할 경계상의 모든 index rectangle들을 양분함으로써 이들이 서로 중첩되지 않도록 한다. 이러한 중복된 객체 엔트리는 결국 노드의 수를 증가시켜 검색영역이 커지는 경우 성능에 치명적인 영향을 미치게 된다.  $R^+$ -tree가 면적을 갖는 데이터의 구성에 있어서는 중복된 데이터에 의한 overhead 때문에 검색면적이 늘어남에 따른 검색성능의 지연이 가장 큰 단점이지만 검색 영역이 아주 작은 경우 예를 들어 point 검색의 경우에 어떠한 공간 색인구조도  $R^+$ -트리 보다 우수한 성능을 보일 수 없는 이유가 바로 여기에 있는 것이다. 단층인식 시스템에서는  $R^+$ -tree가 각 지층의 구성물의 시작점과 끝점의 포인트데이터로 구성되므로  $R^+$ -tree의 장점을 최대한 활용할 수 있다.  $R^+$ -tree의 구현에 있어서 index rectangle들은 MBR(minimum bounding rectangle)을 사용하지 않는다. 그 이유는

MBR을 사용하여 구현 할 때  $R^*$ -tree의 알고리즘이 구현하기 거의 불가능하므로 k-d-b tree에서와 마찬가지로 공간분할을 하여 index rectangle사이에 빈 간격이 생기지 않도록 구성한다.

### 3. 단층인식기법

본 논문에서 사용된 입력 데이터는 Boonsville 지역에서 얻어진 광범위한 3차원 탄성파 데이터로서 그림 2는 3차원 데이터중의 하나의 수직 2차원 단면 데이터 중의 일부만을 보여주고 있다. 실제로 얻어진 데이터는 광범위한 지역에 걸친 3차원의 데이터로 구성되어있다. 그림 2에서의 데이터는 래스터 형태로서 351 X 70 개의 셀로서 구성되어있다. 각 셀은 0 또는 1의 값을 갖는데, 해당 셀의 위치에서 탄성파의 강도가 주어진 threshold이상인 값을 가질 때 1로서 표현된 것이다.

#### 3.1 지층 object의 검색

단층인식의 첫 번째 단계로서 주어진 입력 데이터(그림 2)로부터 지층을 구성하는 object를 인식하여 파악된 object들을 리스트로 구성하는 것이다. 즉, 연속된 1로 구성된 셀 들을 파악하여 그것을 하나의 지하 지층 object로서 인식하는 것이다. 각각의 각 지층 object는 말단점 즉, 시작점과 종료점으로 표현된다. 검색된 시작점과 종료점의 좌표값은 파일에 쓰여지고 이 좌표값을 활용하여 공간자료구조를 구성하게 된다. 각 셀에 flag값을 사용하여 이미 검색되어 하나의 object의 구성원으로 사용된 셀을 중복하여 방문하면서 검색하는 비효율을 방지한다. 셀 값으로 구성된 배열을 행 위주(row major) 방식으로 읽으면서 0이 아니면서 flag값이 0인 셀을 만나면 해당 셀의 flag를 1로 설정하고 그 셀의 위치에서 이동 가능한 6 방향으로 검색을 한다. 그림 3(a)에서 셀 A의 flag이 0인 경우 A의 연속성을 검사하기 위한 방향은 모두 6가지이다. 상하 방향으로 이동할 필요가 없는 이유는 그림 3(b)에서 보여진다. 그림 3(b)는 그림 1에서 보는 바와 같은 형태에 threshold 이상의 값을 가진 모든 셀을 1로 설정한 상태이다. 이러한 그림은 thinning 과정을 통해 그림 2과 같이 처리되어 상하로 여러 개의 1로 설정된 셀들은 하나의 대표 셀로서 표현된다. 따라서 그림 3(b)의 타원에서 보는 것과 같은 형태의 패턴은 제외된다. 따라서 하나의 셀에서 이동 가능한 경로는 모두 6가지가 된다. 셀 A에서 이동방향을 왼쪽 또는 오른쪽으로 분리하여 3가지 방향

으로 한쪽으로만 진행시킬 경우 검색과정에서 edge가 검색되지 못하는 경우가 발생할 수 있다. 그림 4에서 셀 x를 방문한 다음 오른쪽 검색을 위해 3가지 방향으로 진행시킬 경우 오른쪽 끝점인 c를 발견할 수 있으나 진행방향과 반대편으로 구성된 edge b를 검색할 수 없다. 하나의 지층 object가 여러 갈래의 경로를 취할 수 있는 가능성이 있기 때문이다. 따라서 모든 방문한 셀에서 이동 가능한 모든 6방향으로 검색 알고리즘을 재귀적으로 수행하여 모든 말단점을 검색하여야 한다. Leftmost 셀의 좌표값과 rightmost 셀의 좌표값은 각각의 list로 구성이 되고 이 리스트는 각각의 파일로 생성된다.

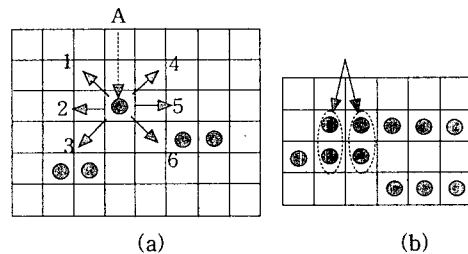


그림 3. 셀의 이동경로

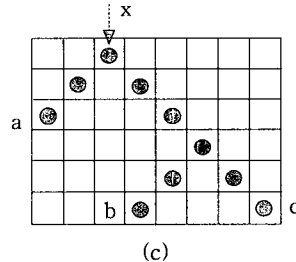


그림 4. object 인식 경로

그림 4에서 셀 x가 최초로 방문되었을 때 왼쪽과 오른쪽으로 이동하면서 말단 셀(a, b, c)을 발견하는 알고리즘은 다음과 같다.

#### Cell Search()

1. 해당 셀의 flag값을 1로 설정한다
2. 해당 셀의 upper left 셀(그림 3(a)의 1번 경로)의 좌표를 생성
3. if 생성된 셀의 위치가 평면의 경계 내에 존재하고  
if 셀의 flag값이 0이면  
flag값을 1로 설정하고 Cell Search()를 재귀적으로 호출한다  
else 이전 셀을 말단 셀로서 list에 삽입

else 이전 셀을 말단 셀로서 list에 삽입

4. 해당 셀의 left 셀(그림 3(a)의 2번 경로)과 left lower 셀(그림 3(a)의 3번 경로), upper right셀(그림 3(a)의 4번 경로), right 셀(그림 3(a)의 5번 경로), right lower셀(그림 3(a)의 6번 경로) 각각의 좌표를 생성하고 위의 3번 과정을 각 경우에 대하여 수행한다

List에 말단 셀의 정보를 저장할 때 해당 셀의 검색 경로의 방향정보와 object ID도 함께 저장한다. 검색 경로정보는 왼쪽 방향의 검색 경로를 가진 셀의 경우 검색 시 동일한 왼쪽 방향의 셀과는 단층을 이룰 수 없으므로 단층여부를 검색할 필요가 없다. 또한 동일한 object에 소속된 말단 cell들은 서로 단층을 이루는 쌍을 구성할 수 없기 때문에 셀의 object ID를 사용하여 불필요한 검색을 피할 수 있다.

### 3.2 지층 object의 구성과 단층 검색

발견된 말단 셀들은 검색 경로에 따라 leftmost 셀과 rightmost 셀로서 각각의 리스트로 구성되어 있다. 이들 리스트에 대해 둘 중 하나를 선택하여 R<sup>+</sup>-tree를 구성한다. 예를 들어 rightmost 셀을 R<sup>+</sup>-tree로 구성하는 경우를 가정하여 보자. 셀의 위치는 단순한 점으로서 x, y 두 개의 좌표로서 구성될 수 있다. R<sup>+</sup>-tree의 구성이 매우 간단해지고 R<sup>+</sup>-tree의 공간분할이 이러한 포인트 데이터를 구성하는데 최적의 성능을 발휘할 수 있다. 구성된 R<sup>+</sup>-tree에 대해 leftmost 셀이 저장된 리스트의 각 셀에 대해 가장 근접한 셀을 검색한다. 검색방식을 주어진 leftmost 셀의 위치에서 가장 인접한 이웃(nearest neighbor) 셀(rightmost)을 발견하는 방식이다. 최초의 검색범위 크기는 전체 object의 평균길이를 사용하여 설정하고 최소검색에서 검색범위에 rightmost 셀이 발견되지 않을 경우 검색범위를 확장시켜 검색을 진행한다. 검색범위의 확장범위도 object의 평균길이와 축적된 실험의 결과에서 얻어진 값을 사용하여 조정한다. 또한 검색범위가 일정크기 이상으로 확대되어도 검색에 실패하는 경우 해당 지층과 match되는 지층이 존재하지 않는 것으로 한다. 검색범위에서 발견된 셀 중에서 object ID를 비교하여 동일한 ID를 가진 object를 제거한다. 이것은 동일한 지층 object의 시작과 종료좌표가 하나의 단층구간으로 인식되는 것을 방지하기 위한 것이다. 발견된 object pair는 하나의 단층구간을 구성하는 후보로서 list에 저장하여 최종적인 검색이 종료된 후에 Arc/view[5]와 같은 시각화 도구를

사용하여 해당구간을 표시하여 지질 전문가의 최종적인 해석을 위한 전처리 과정이 된다.

## 4. 결론

최근에 지리정보시스템과 같은 정보의 시각화는 컴퓨터 기술의 중요한 한 부분으로서 그 응용분야의 영역이 점차 커지고 있다. 시각정보는 일상적인 업무처리에서 발생하는 의사결정에 있어 중요한 부분을 차지한다[3]. 예를 들어, 통신, 국방, 환경, 도시계획, 토목, 자원관리, 교통, 통계, ...등 거의 모든 현대의 응용분야가 그 범주에 들어갈 수 있다[1, 4]. 그러나 지하정보에 대한 연구는 지하정보의 비가시적인 특성상 아직은 초기단계라고 할 수 있다. 특히 지하 지층 분포를 체계적으로 분석하기 위한 연구는 지하 광물자원에 대한 정보 또는 지상의 구조물에 대한 안전성 평가에 대단히 중요한 역할을 한다. 본 논문은 지하 단층을 공간자료구조를 활용하여 인식하는 시스템 개발 방법을 제안하고 이와 관련된 사항을 기술하였다. 이러한 시스템의 개발은 지질 전문가의 지질모델 작성 및 해석을 할 수 있는 전처리 시스템으로서 중요한 역할을 할 것이다.

## [참고문헌]

- [1] Tor Bernhardsen, Geographic Information System, Norwegian Mapping Authority, 1992.
- [2]Guttman, A., R-trees: A Dynamic Index Structure for Spatial Searching, Proc. of the ACM SIGMOD, pp. 47-57, 1984.
- [3] Understanding GIS, The Arc/Info Method, 1994.
- [4] Michael F Goodchild, Introduction to GIS, National Center for Geographic Information and Analysis Univ. of California, 1991.
- [5]Hoel, E. G., and Samet, H., A Qualitative Comparison Study of Data Structures for Large Segment Databases, ACM SIGMOD, pp. 205-214, 1992.
- [6]Sellis, T., Roussopoulos, T. and Faloutsos, C., R<sup>+</sup>-tree: A Dynamic Index for Multi-dimensional objects, Proc. of the 13th VLBD Conference, pp. 507-518, 1987.