

# 확장된 시공간 데이터 집합 생성기<sup>†</sup>

이선준\*, 김상호\*, 류근호\*, 이성호\*\*

\*충북대학교 데이터베이스 연구실

\*\*한국전자통신연구원

e-mail:{sunjun, shkim, khryu}@dbl.chungbuk.ac.kr

## The Extended Generator of Spatiotemporal Datasets

Sun Jun Lee\*, Sang Ho Kim\*, Keun Ho Ryu\*, Seong Ho Lee\*\*

\*Database Laboratory, Chungbuk National University

\*\*Electronics and Telecommunications Research Institute

### 요약

시공간 접근 방법들을 위한 효율적인 성능평가 환경은 최소한 다음과 같은 모듈들을 포함해야 한다. 종합적인 데이터집합의 생성, 데이터집합의 저장, 접근 구조들의 수집과 실험, 실험적인 결과의 시각화 등이다. 데이터집합 저장 모듈에 초점을 맞추어서 다양한 실제 세계 시나리오를 실험하기 위한 종합적인 데이터의 생성이 요구된다. 과거의 여러 알고리즘들은 작업공간에서 미리 분배된 정적인 공간데이터를 생성하기 위하여 구현되어져왔다. 하지만 시간에 따라 변화하는 공간객체인 시공간 데이터를 생성하기에는 어렵다. 이 논문에서 시공간 데이터 타입의 데이터 생성기에서 고려하여야 할 매개변수들에 대하여 논의한다. "Generate\_Spatio\_Temporal\_Data"라는 알고리즘은 움직이는 점 또는 사각형데이터를 생성하고 거래시간과 유효시간을 구별하지 않았으며 시간 점만을 표현하였다. 이 논문에서는 정확한 시간적 개념을 표현하기 위하여 거래시간과 유효시간간격 모두를 지원하는 데이터 생성기에 관하여 논의하기 위한 알고리즘을 제시하고 실제적인 데이터집합 생성을 위한 매개변수들을 나타낸다.

### 1. 서론

공간데이터베이스와 GIS에서 현재 연구중인 분야는 시간에 따라 위치, 형태, 크기가 변하는 객체를 포함하는 현실적이고 지리적인 응용들의 정확한 모델링을 수반한다. 마찬가지로 시공간데이터베이스시스템의 주요한 목표도 실제계의 정확한 모델링이다. 즉 실제계의 예로서 화재, 태풍의 경로 추적, 비행기 모의실험, 기상예보등을 포함한다.

데이터베이스시스템은 이러한 응용들을 지원하고 효율적인 모델링을 위해 확장되어야한다. 이러한 목표를 위하여 최근의 연구 노력, 결과들은 다음과 같은 것들을 목표로 하고 있다. 시간에 따라 변화하는 공간객체들에 대한 모델링과 질의처리, 접근방법들의 설계, 적절한 구조들과 시스템의 구현이다. 시공간 데이터베이스는 공간상에 존재하는 객체의 공간적 요소와 비공간적 요소를 표현하고 이들에 대한 다양한 연산을 제공하며, 시간의 흐름에 따라 변화하는 객체에 대한 이력을 관리해 준다[3]. 시공간 데이터는 시공간데이터베이스에서 관리하는 주요 데이터로서 매우 복잡하고 대용량이라는 특성을 갖는다. 따라서 이러한 데이터를 효과적으

로 저장 관리하기 위해서는 시공간 데이터에 대한 효율적인 접근 방법에 대한 연구가 필요하다. 이 작업은 시공간 데이터의 질의에 대한 효율적인 인덱스 기술과 빠른 접근 방법이 요구된다. 기존의 공간접근방법들의 확장 또는 새로운 방법이 있을 수 있다. 하지만 모든 제안들은 현실적이고 종합적인 광범위한 실험 하에서 평가되어야 한다. 전반적으로 저장공간, 생성시간, 공간과 시간과 시공간 질의에 대한 응답시간측면에서 제안된 접근방법들 사이의 일관된 성능 비교가 부족하다. 시공간 접근 방법들에 대한 성능 평가 환경의 일반적인 구조는 종합적인 데이터와 질의 집합들을 생성하는 모듈, 실제 데이터집합들의 저장, 실험적인 목적의 접근 구조들의 수집, 실험적인 결과들의 데이터베이스, 설명 목적을 위한 데이터집합들과 구조들을 보여주기 위한 시각적인 불등을 포함한다.

성능평가 또는 실험 관리를 위한 독립적인 플랫폼에 대한 필요는 이미 과거에 논의되어왔다. 확장된 관련 연구가 전통적인 데이터베이스 성능평가와 데이터 생성기에서 발견되어지지만 시공간데이터베이스 분야에서는 매우 제한적이다. 따라서 이 논문에서는 명확한 시간개념을 가지는 시공간 데이터집합들을 생성하기 위하여 고려해야하는 매개변수들을 검토한다. 그리고 사용자 요구 측면에서의 다양한 시나리오

<sup>†</sup> 이 연구는 한국과학재단 RRC(청주대 정보통신연구센터)와 ETRI의 연구비 지원으로 수행되었음.

들을 실험하기 위한 데이터집합의 튜플 구조를 제안한다.

2장에서는 시공간 인덱스의 명세에 대하여 논의한다. 3장에서 기존의 시공간인덱스에 살펴보고 4장에서는 확장된 GSTD에 대한 논의와 매개변수 변경을 통한 여러 시나리오와 대응되는 결과를 나타낸다. 5장에서는 향후연구와 결론으로 끝을 맺는다.

## 2. 시공간인덱스 명세

[6]은 지원되는 데이터 타입들과 데이터집합들, 인덱스 구성, 질의 처리 측면에서 효율적인 시공간 접근 방법들을 평가하고 설계할 때 고려되어야 하는 명세들의 목록에 대해 논의한다. 특히, [6]에서 논의된 첫 번째 측면은 지원되는 공간적인 데이터타입, 지원되는 시간적인 데이터타입, 데이터집합의 이동성(mobility)등이다. [8]에서 "degenerate"라는 용어는 객체 인스턴스의 유효시간은 그것들의 거래시간과 동일하다는 특성을 언급한다. 즉 객체는 데이터베이스에 존재하는 동안 유효하다는 것이다. 하지만 두 시간적 개념은 구별되어야한다. 유효시간은 모델화된 세계에서 사실이 참일 때의 시간이며 거래시간은 데이터베이스에서 사실이 현실일 때의 시간을 말한다. 유효시간은 자유롭게 수정될 수 있지만 거래시간은 수정될 수 없다. 그러므로 이 논문에서는 위 시간적 개념의 구별을 가지고 시공간 데이터를 생성하였다.

종합적인 시공간 데이터집합들을 생성하는데 있어서 기본적인 이유는 공간 객체들의 전개를 다룰 수 있는 매개변수들의 완전한 집합의 정의이다. 이러한 목표를 위하여 다음과 같은 세가지 연산들이 있다.

- 객체 인스턴스의 간격(duration), 연속적인 인스턴스 사이의 타임스탬프의 변화
- 객체의 이동(shift), 공간적인 위치의 변화
- 객체의 크기변화(resizing), 객체의 크기 변화(점이 아닌 객체에서 이용)

세계의 매개변수 interval,  $\Delta center[]$ ,  $\Delta extent[]$ 를 이용하여 객체의 new\_timestamp(숫자값), new\_spacetime\_center(2차원적인 점), new\_spacetime\_extent[(배열)를 계산할 수 있다. 그리고 성능평가 환경은 광범위한 초기 데이터 분배를 지원해야한다. 위 매개변수들에서 다른 분배들의 주의 깊은 사용을 통해서 여러 가지 흥미로운 시나리오들을 모의 실험할 수 있다.

## 3. 기존의 시공간인덱스

여러 접근방법들이 시간적 측면에 대한 고려 없이 공간데이터를 지원하기 위해 제안되어져왔다. 이러한 방법들은 점들과 사각형들, 임의의 형태를 가진 객체들과 같은 지리적 객체들을 다룰 수 있다. 전체적인 관련연구는 [5]에서 볼 수 있다. 반면에 시간적인 접근 방법들은 유효시간과 거래시간을 인덱스하기 위하여 제안되어져 왔다. 시간 접근 방법들을 위한 연구는 [9]에서 볼 수 있다. 객체의 공간과 시간 모두를 고려한 여러 기존의 시공간인덱스에 관하여 살펴본다. 특히, 3D R-트리, HR-트리, GR-트리는 R-트리에 기반을 두고 있다. 반면에 Overlapping Linear Quadtree는 Quadtree구조에 기반을 두고 있다. 3D R-트리[2]는 시간을 이차원 공간 내의 또 다른 차원으로 간주하여 이차원 영역들을 시간 축을 고려한 3차원 MBR형태로 표현한다. 일반적인 R-트리를 이용하여 멀티미디어 데이터의 색인을 위해 제안되었으며 유효시간차원만을 색인한다.

HR-트리[4]는 원래의 트리는 유지하고 상태가 변한 루트와 가지를 대체함으로써 R-트리의 현재와 과거를 유지하는

것이다. 상태가 변하지 않는 가지는 복제되지 않는다.

MV3R-트리[10]는 거래시간데이터베이스에서 1차원 데이터의 변화를 다루는 MVB-트리의 개념을 이용한 것으로 MVR-트리와 단말노드들로 구성되는 보조적인 3DR-트리를 결합하여 타임스탬프질의와 간격질의를 효율적으로 처리할 수 있다.

GR-트리[7]는 시간 데이터베이스에서 유효시간과 거래시간을 모두 다루며, now-relative data를 다루기 위해 R-트리를 확장한 색인 구조이다.

대부분의 제안된 시공간 접근 구조들은 R-트리에 기반을 둔다. 우리도 그러한 구조에 초점을 맞춘다. 위의 제안된 접근들 사이에서 3D R-트리는 종합적이고 균일한 데이터집합을 사용하여 구현되고 실험되었을 뿐만 아니라 여러 시간, 공간, 시공간 연산들을 위한 검색비용이 측정되었다.

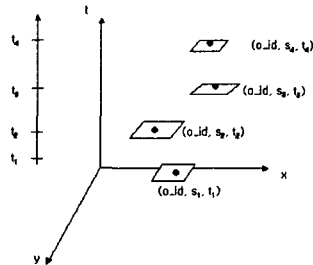
## 4. 확장된 GSTD 알고리즘

[8]에서 시간에 따라 변하는 점, 사각형 객체들을 생성하기 위한 Generate\_Spatio\_Temporal\_Data(GSTD)라는 알고리즘을 제안하였다. 각 객체 o\_id에서 GSTD는 (o\_id, t, p, p\_u)형식의 튜플들을 생성한다. 여기서 거래시간과 유효시간의 개념을 구별하지 않는 시간점을 나타내는 튜플들을 생성한다(그림1). 하지만 이 논문에서는 정확한 시간개념을 나타내기 위하여 각 객체 o\_id에서 (o\_id, vt1, vt2, tt1, tt2, p, p\_u)형식의 튜플들을 생성하는 확장된 GSTD알고리즘을 제안한다(그림2). vt1, vt2는 유효시간간격을 나타내고 tt1, tt2는 거래시간간격을 나타내는 타임스탬프이다. p1(p\_u)는 공간스탬프의 낮은(높은) 좌표점이다. 확장된 GSTD알고리즘은 그림2에서 설명된다. 대부분의 함수는 원래 GSTD알고리즘의 함수를 이용한다.

### 4.1 알고리즘에 대한 설명

그림2는 확장된 GSTD가 생성하는 객체의 인스턴스에 대한 예를 나타낸다. (a)는 객체 o의 위치변화에 따른 거래시간간격을 나타내며 (b)는 객체 o의 위치변화에 따른 유효시간간격을 나타낸다. 그림3에서 (a)는 객체 o가 s1위치에서 tt1, tt2의 거래시간간격과 vt1, vt2의 유효시간간격을 가지는 것을 나타내며 (b)는 객체 o가 s2위치에서 tt2, tt3의 거래시간간격과 vt1', vt2'의 유효시간간격을 나타낸다.

그림4에 표현된 알고리즘은 다음과 같다. 확장된 GSTD는 입력으로서 많은 사용자 정의 매개변수들을 받는다. N과 D는 각각 데이터집합의 초기 카디널리티와 밀도, starting\_id는 객체들의 초기 id, numsnapshots는 작업공간의 시간적 분해, min\_t와 max\_t는 interval매개변수의 영역, min\_c[]와 max\_c[]는 center[]매개변수의 영역, min\_ext[]와 max\_ext[]는 extent[]매개변수의 영역이다. 다음과 같은 처리에 따라 각 객체들의 여러 튜플들을 생성한다.



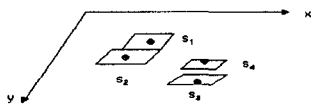
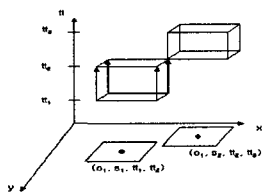
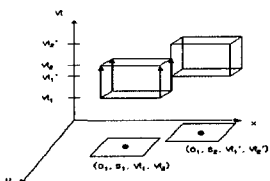


그림1. GSTD에서의 시간에 따라 변화하는 객체의 예  
 각 객체들은 초기에 active하고 각각에 대하여 새로운 인스턴스들은 그것들의 시간들이  $vt_2$ ,  $tt_2 < 1$  일 때까지 생성된다. 모든 객체들이 inactive일 때 알고리즘은 끝난다. 초기화 부분에서 모든 객체들은 초기화되고 그것들의 중앙 점들은  $distr\_init()$ 분배에 기반 하여 작업공간에서 임의적으로 분배된다. 주된 루프 부분에서 객체의 각 새로운 인스턴스는 기존의 것과 세 개의 매개변수 대한 함수를 이용하여 생성된다. 유효하지 않은 인스턴스들은 세가지 선택적인 방법들(*radar*, *adjustment*, *toroid*)을 통하여 조작된다. 생성되어 질 새로운 인스턴스를 위하여 *interval*, *center*, *extent*값들은  $RNG(distr(), min, max)$ 루틴을 호출함으로써 계산된다.



(a) 거래 시간 측면



(b) 유효시간 측면

그림2. 확장된 GSTD가 생성하는 객체의 예  
 다음의 미리 정의된 *dist* 따라 *min*과 *max*사이의 임의의 수들을 생성한다. *print\_instance*함수는 객체의 현재 인스턴스는 *next\_snapshot*에서 값이 같거나 큰 타임스탬프값을 갖는지를 체크한다. 그렇다면 현재 인스턴스전의 인스턴스의 좌표들이 적절한 타임스탬프를 사용하여 출력된다.

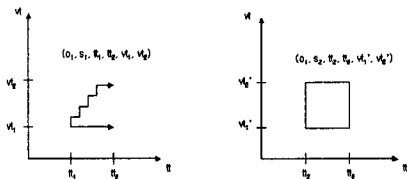


그림3. 객체의 위치변화에 따른 시간적 측면만 표현한 객체의 예  
 추가적으로 *next\_snapshot* 변수의 값은 적절하게 조절되고 그렇지 않다면 현재 인스턴스는 출력되지 않는다. 작업공간의 밖에 좌표가 위치할 수 있다. GSTD는 이러한 유효하지 않은 인스턴스들을 세가지 선택적인 접근방법들(*radar*, *adjustment*, *toroid*)중 하나에 따라 다룬다.

#### 4.2 일반적인 데이터집합 예들

이전에 언급한대로 시공간 데이터집합들의 실제적인 예들은 인간, 동물, 자동차의 경로추적을 포함한다. 그리고 이것은 GPS, 비행 시뮬레이션, 기상예보, 화재나 태풍 추적에 의해 검출되기도 한다. 예를 들면 GPS에 의해 자동차 이동을 검출하고 데이터베이스에 전체적인 경로를 저장하는 것은 전형적인 일상생활에서의 예이다.

```

Input: values N, starting_id, numsnapshots, D, min_t, max_t
arrays min_c[], max_c[], min_ext[], max_ext[]
distributions distr_init(), distr_t(), distr_c(), distr_ext()
output: instance(id, vt1, vt2, tt1, tt2, lower_left_point,
upper_right_point), validity_flag

begin
for each id in range [starting_id, N + starting_id] do
Set vt1=vt2=tt1=tt2=0, center[] = RNG(distr_init(), 0, 1),
extent[] = extent(N,D)
Set active = TRUE
end for
Set step = 1 / numsnapshots
for each id in range [starting_id, N + starting_id] do
Set next_snapshot = step
while active do
Set intervalvalid = RNG(distr_t(), min_t, max_t)
Set intervaltransaction = RNG(distr_t(), min_t, max_t)
Set Δcenter[] = RNG(distr_c(), min_c[], max_c[])
Set Δextent[] = RNG(distr_ext(), min_ext[], max_ext[])
Set old_instance = instance
update_instance(instance)
if ((vt2 > 1) && (tt2 > 1))then
active = FALSE
print_instance(old_instance, current[i], next_snapshot)
else
Set validity_flag = valid(instance)
if validity_flag = FALSE and approach != 'radar' then
adjust_coords(instance, approach)
end if
if((vt1 > next_snapshot) && (vt2 > next_snapshot)
&& (tt1 > next_snapshot) && (tt2 > next_snapshot))
then
print_instance(old_instance, current[i], next_snapshot)
end if
end while
end for
end.
    
```

그림4. 확장된 GSTD 알고리즘

하지만 다른 이동 시나리오들은 효율적인 구조가 평가되어야 하는 다른 데이터집합에 대응된다. 치우친 방향 대 임의의 방향, 느린 이동 대 빠른 이동은 전체적으로 다른 응용들을 생성하는 몇가지 매개변수들이다.

몇가지 이러한 시나리오들을 실험평가하기 위하여 확장된 GSTD에 의해 생성된 점과 사각형 객체들로 구성할 수 있다. 세가지 접근방법(*radar*, *toroid*, *adjustment*), *N*(투플의 개수), *D*(분포정도), *numsnapshots*(스냅샷의 개수)등의 여러 매개변수들의 조합을 통해서 다양한 데이터집합을 생성할 수 있다. 위치변화와 크기변화의 임의성은 *center[]*와 *extent[]*를 사용하는 *Uniform(min, max)* 분배에 의해 보장된다. *center[]*의 *min*과 *max* 값들을 증가시킴으로서 사용자는 'faster'객체를 얻을 수 있다. 그래서 데이터 집합의 속도는 *meta-information*으로서 간주될 수 있다. 이것은 기본적인 매개변수들로부터 유도될 수 있다.

선택적으로 만일 사용자의 응용이 두가지 또는 그 이상의 시나리오 결합이 필요하다면 예를 들어, 어떤 방향으로 움직이는 약간의 MBR들의 집단과 나머지는 고정되어 있을 수 있다. 두 별개의 시나리오로 두 *starting\_id* 입력매개변수들을 적절하게 설정함으로써 생성되어질 수 있고 결합된다. 요컨대 매개변수들을 조정함으로써 필요한 시나리오를 산출할 수 있다. 확장된 GSTD로 생성된 데이터는 시간에

다른 위치변화를 중요시하는 차량추적시스템보다는 정보에 대한 유효성을 중요시하는 추가변동시스템이나 토지정보시스템등을 실험평가하기에 적합하다.

## 5. 결론

시공간데이터베이스시스템은 시공간 데이터에 대한 적절한 인덱싱 기술들이 필요하며 시공간 인덱싱들의 공정한 성능 비교를 제공하기 위한 성능평가 환경의 구현과 설계가 제공 되어야한다. 또한 다양하고 적절한 종합적인 실제 시공간 데이터집합들을 수집할 필요가 있다.

우리는 GSTD에 정확한 시간적 개념을 지원하기 위하여 거래시간과 유효시간 간격을 가진 튜플을 생성하기 위한 확장된 GSTD생성기를 제안하였다. 이러한 연구는 시공간데이터베이스시스템의 목적들을 위한 접근방법들을 설계, 구현, 평가하기 위한 상호적인 지원을 향한 것으로 [9]에서 시작하여 계속되고 있다. 더 다양한 시나리오를 지원하기 위해 많은 실험을 통해 또 다른 매개변수를 발견하는 것과 여러 시공간인덱스에 대하여 확장된 GSTD로 생성된 데이터집합을 가지고 실험평가를 준비하고 있다.

## 참고문헌

- [1]Guttman, A, R-trees: A dynamic index structure for spatial searching, In Proc. ACM SIGMOD Int. Conf. on Management of Data, 1984, pp.47-54.
- [2]Y.Theodoridis, M.Vazirgiannis, T.Sellis, Spatio-Temporal Indexing for Large Multimedia Applications, Submitted to the 3rd IEEE International Conference on Multimedia Computing and Systems(ICMCS 1996), pp.1-23.
- [3]D.H.Kim, K.W.Nam, and K.H.Ryu, "Spatiotemporal Data Model," Proceedings of The 23th KISS Fall Conference, pp.201-204, 1996.
- [4]Mario.A.Nascimento, Jefferson.R.O.Silva, TOWARDS HISTORICAL R-TREES , ACM 1998, pp.235-240.
- [5]Volker Gaede, Oliver Gunther, Multidimensional Access Methods. ACM Computing Surveys, 30(2), 1998, pp.170-231.
- [6]Y. Theodoridis, T. Sellis, A. Papadopoulos, Y. Manolopoulos, "Specifications for Efficient Indexing in Spatiotemporal Databases", Proceedings of the 6th ACM International Workshop on Geographical Information Systems(ACM-GIS), 1998.
- [7]Simonas Saltenis, Christian S.Jensen, R-Tree Based Indexing of General Spatio-Temporal Data, CHOROCHRONOS TECHNICAL REPORT CH-99-18, December 1999.
- [8]Y.Theodoridis, J.R.O.Silva, M.A.Nascimento, ON THE GENERATION OF SPATIOTEMPORAL DATASETS, CHOROCHRONOS TECHNICAL REPORT CH-99-01, JANUARY 1999.
- [9]Betty Salzberg, Vassilis J. Tsotras, A Comparison of Access Methods for Time Evolving Data, Science and Technology Foundation as part of its Center for Advanced Technology program.
- [10]Yufei Tao, Dimitris Papadias, The MV3R-Tree: A Spatio-Temporal Access Method for Timestamp and Interval Queries, 27th VLDB 2001