

이동체 데이터베이스의 공간 색인 성능평가

이주형^o, 김진덕^{**}, 전봉기^{***}, 홍봉희^{***}

^o부산대학교 GIS학과, ^{**}동의대학교 컴퓨터공학과, ^{***}부산대학교 컴퓨터공학과
zenith2@hyowon.cc.pusan.ac.kr, jdk@dongeui.ac.kr,
{bgjun,bhhong}@hyowon.pusan.ac.kr

Performance Evaluation of Spatial Indices for Moving Object Database

Ju-Hyoung Lee^o, Jin-Deog Kim^{**}, Bong-Gi Jun^{***}, Bong-Hee Hong^{***}

^oDept. of GIS, Pusan National University

^{**}Dept. of Computer Engineering, Dongeui University

^{***}Dept. of Computer Engineering, Pusan National University

요 약

이동체는 끊임 없이 위치 정보를 변경하는 특징을 가지고 있다. 빈번한 데이터 변경이 발생하는 이동체에 대한 효율적인 색인 기법의 대한 연구가 필요하다.

이 논문에서는 이동체의 공간 데이터 표현 방법과 색인 변경 정책을 제안한다. 그리고 제안한 색인 변경 정책에 따른 공간 색인별 성능 평가와 구축된 공간 색인의 영역 질의 처리에 대한 성능 평가를 위한 실험 평가기를 설계한다. 실험을 통해 이동체 데이터베이스에서 효율적으로 사용 가능한 색인을 도출해 낼 수 있으며, 실험 결과물의 하나인 성능 평가기를 사용하여 향후 개발할 이동체를 위한 새로운 색인에 대한 성능 평가도 수행할 수 있다.

1. 서론

최근 들어 GPS의 활용도가 높아지고, 무선 데이터 통신의 기술이 발달함에 따라 이동체의 위치 정보를 수집하기가 용이해졌다. 수집된 이동체의 데이터는 효율적인 데이터 관리를 위해서 데이터베이스에 저장된다. 그러나 끊임없이 이동하며 위치 정보를 변경하는 이동체를 전통적인 공간 데이터베이스 시스템에 저장 및 관리할 경우 심각한 시스템 성능 저하를 초래하게 된다.

성능 저하의 원인으로는 첫째, 전통적인 공간 데이터베이스 시스템에 저장되는 데이터는 정적 데이터이며, 저장 이후 영시적인 변경이 발생하지 않는 이상 지속적(Constant)이라는 점이다[1]. 둘째, 전통적인 공간 데이터베이스 시스템에서는 사용하는 공간 색인은 정적 데이터의 효율적인 검색을 위주로 고안된 것이기 때문에 이동체의 고유한 특성을 반영하지 못하고 있다. 따라서 빈번한 데이터 변경이 발생하는 이동체 데이터베이스에 대한 효율적인 색인 기법의 대한 실험이 필요하다.

이 논문에서는 이동체를 공간 데이터베이스에 효율적으로 저장하기 위한 모델을 제안한다. 저장된 이동체의 특성을 반영하는 효율적인 변경 정책을 제시한다. 또한 변경 정책의 성능을 측정하기 위하여 다음과 같이 두 종류의 실험을 한다. 첫째, 색인 변경

정책에 따른 공간 색인별 성능 측정과 둘째, 구축된 공간 색인에 대한 영역 질의 처리에 대한 성능 평가이다.

논문을 통하여 이동체에서 사용할 색인의 설계 및 사용에서 성능을 위해 고려해 할 인자를 밝혀 낼 수 있으며, 논문의 결과인 성능 평가를 사용하여 향후 개발할 이동체를 위한 새로운 색인의 성능 평가도 수행할 수 있다.

이 논문의 구성은 2장에서 이동체 저장을 위한 모델과 공간 색인 변경 기법에 대해서 알아본다. 3장에서는 제안한 기법의 성능 평가를 위해 사용할 성능 평가기의 구조와 실험 평가 인자 및 실험에 대한 구체적인 내용을 기술한다. 결론 및 향후 연구는 4장에서 기술한다.

2. 이동체를 위한 공간 색인 방법

이 장에서는 이동객체를 위한 데이터 표현 방법과 변경이 잦은 이동 객체를 효과적으로 색인하기 위한 공간 색인 방법을 제안한다.

2.1 이동체 표현방법

기존 연구에서는 이동체를 대부분 시간의 함수[2]로 표현하고 있다. 이러한 표현 방법은 이동체 추적을 목적으로 하는 것이며,

현재 이동체의 위치 검색에 초점을 두는 본 실험에는 적합하지 않다.

이 논문에서 이동체(Mob)를 유일한 식별자(Mobj.id), 위치 보고 시간(Mobj.time) 그리고 위치 보고 시간에 따라 갱신되는 위치 정보(Mobj.loc)를 가지는 객체라고 정의한다. 데이터베이스에 저장 될 때에는 이동체는 가장 최근에 보고된 위치 데이터만 저장된다고 본다. 예를 들어, 그림1]에서 현재 시간에서 이동체A는 (Mobj.a, Mobj.loc₃, t₃)의 값으로 공간 데이터베이스에 저장된다.

위치 정보 수집에 장애가 생기거나, 무선 데이터 통신 상의 장애를 고려해 보았을 때, 위치 갱신 시간 간격 사이의 이동체 위치 값들은 유효하다고 가정한다. 즉, 그림1]에서 시간 t₀에서 위치 변경 보고 후 시간 t₁의 변경이 발생하기 전까지 이동체A는 (loc₀, t₀) 값을 가진다.

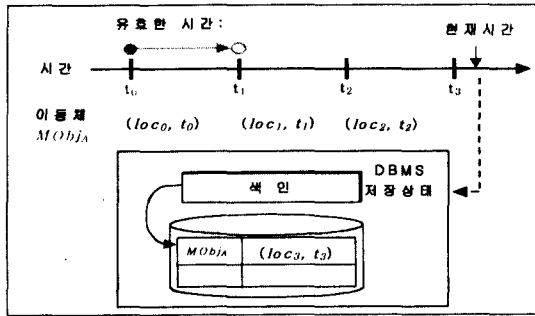


그림 1 이동체 표현 방법

2.2 이동체의 색인을 위한 색인 변경 정책

일반적으로 공간 데이터베이스의 색인은 객체가 존재하는 공간을 일정 간격으로 분할 하는 방법(Space Decomposition)과 객체의 최소 영역을 이용하여 계층적으로 트리를 구성하는 방법(Object Bounding)으로 나누어 볼 수가 있다.[3]

이 실험에서는 각 방법에서 대표성을 가지는 색인인 Grid File[3]과 R+-tree[4]를 실험 대상 색인으로 한다.

이동체의 데이터는 일정 시간 간격 사이에 최소한 한번은 데이터베이스를 변경한다. 변경이 발생하면 데이터베이스에 저장된 데이터뿐만 아니라 색인에도 변경을 발생 시킨다. 데이터베이스에서 색인의 변경은 데이터베이스에 저장된 데이터 값의 변경과는 달리, 기존의 구축된 기존의 색인을 검색, 삭제, 삽입 하는 일련의 연산을 필요로 한다.

R+-tree는 삭제, 삽입 연산에서 발생하는 변경이 상위 노드까지 상위 변경 전파되는 성질[5]을 가지고 있다. 이동체의 위치 보고 간격이 매우 짧다면 변경 발생 즉시 색인에 반영하기 보다는 색인에서 일정 시간 간격을 두고 재구축하는 방법이 효율적일 수 있다. 따라서 이 논문에서 색인 구축 정책을 첫째, 이동체의 데이터 변경 보고 즉시 색인 변경 방법(Immediate Update Policy : IUP)과 둘째, 색인 구축을 일정 시간 간격을 두고 색인을 재구축하는 방법(Deferred Update 또는 Rebuild Policy : RP)이 두 가지 방법을 제안한다.

제안한 방법의 성능 평가 실험에 앞서서, 제안한 두 방법을 이용하여 색인의 변경 후 동일한 질의 성능을 보장하기 위해서는 색인 구축 시간 간격에 대한 정의가 필요하다. 그림 2]에서 보는 것과 같이 재구축 시간 간격(RTI: Rebuild Time Interval)이 보고 즉시 변경 시간 간격(IUTI: immediate Update Time interval)보

다 크다면 동일한 질의에 대해서 서로 다른 결과를 초래하게 된다.

따라서, 실험에서 재구축 시간 간격(RTI)은 수식 (1)과 같이 최소한 보고 즉시 변경 시간 간격(IUTI)의 절반보다 같거나 작아야 동일한 검색 성능을 보장 받을 수 있다.

$$RTI \leq \frac{IUTI}{2} \quad \text{--- (1)}$$

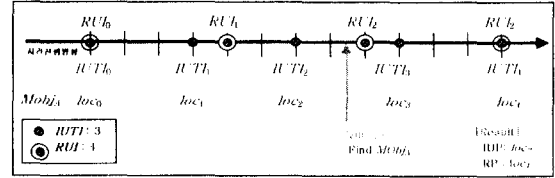


그림 2 색인 구축 시간 간격의 문제점

3. 실험 평가기와 실험 평가요소

이 장에서는 실험에서 사용할 실험 평가기의 전체적인 구조와 각각의 구성 요소들에 대해서 기술한다. 다음으로 실험 인자와 구체적인 실험 시나리오를 기술한다.

3.1 실험 평가기

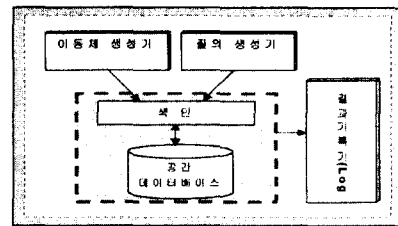


그림 3 실험 평가기 구조

- 이동체 생성기
 - 실험에서 사용할 데이터 집합은 실제 도로 데이터 집합이며, 이동체의 위치 이동 경향을 수학적 분포를 실험 인자로 하여 이동체를 일정 시간 간격으로 생성해 낸다.
 - 실험에서 사용하는 수학적 분포는 정규분포를 사용한다. 또한 특정 지역에 이동체가 집중하는 현상을 표현하기 위해서 집중지역(Spot Area)생성 알고리즘[5]을 사용한다.
- 질의 생성기
 - 제안된 방법으로 변경된 색인에 대한 성능 평가를 공간 질의를 통해서 수행한다. 이때 사용하는 공간질의 종류는 영역 질의이다 질의 영역의 크기는 선형적 증가를 가지도록 한다.
- 결과 기록기
 - 입력된 이동체의 데이터 집합과 질의를 바탕으로 색인의 성능 평가의 수치적 결과를 저장한다. 기록기를 통해서 색인의 성능 평가 결과를 시각적으로 살펴 볼 수 있다.

3.2 실험 인자 및 실험 시나리오

2.2 절에서 제안한 색인 변경 정책의 성능을 평가하기 위해 표 1]과 같이 실험 인자를 설정하고 실험1~4를 수행한다. 모든 실험에서 변경 보고 즉시 색인 변경 방법(IUP)의 동기화율은

[가정 1]을 따른다.

[가정 1]

변경 보고 즉시 색인 변경 방법(IUP)에서 동기화율(m/M, M: 전체 이동체 수, m:정해진 시간 간격 동안 위치 변경 보고를 하는 이동체 수)은 정규분포를 따른다.

실험 인자 1	이동체 수
실험 인자 2	색인 변경 간격
실험 인자 3	색인의 종류 (Grid File, R*-tree)
실험 인자 4	색인 변경 정책 (IUP, RP)
실험 인자 5	Grid Filed의 버퍼 크기
실험 인자 6	R*-tree의 분기 계수(Branch Factor)
실험 인자 7	영역 질의의 영역 크기 비율

표1] 실험 인자

● 실험 인자 1 : 이동체 수

최적의 성능을 나타낼 수 있는 이동체 수를 측정한다.

● 실험 인자 2 : 색인 구축 간격

색인 변경 간격은 변경 즉시 색인 변경 정책(IUP)의 위치 변경 시간 간격(IUT)과 동일하다.

● 실험 인자 3 : 색인의 종류 (Grid File, R*-tree)

● 실험 인자 4 : 색인 변경 정책 (IUP, RP)

● 실험 인자 5 : Grid Filed의 버퍼 크기

Grid File에서 그리드 셀의 크기가 상대적으로 작을 경우 영역 질의에 대해서 우수한 성능을 나타낸다. 그러나 이동체의 이동 거리가 클수록 이동 전/후에 다른 셀에 위치할 확률이 높아지기 때문에 색인의 성능은 떨어지게 된다. 이와 반대로 버퍼 크기를 경우 색인의 변경에서는 비교적 좋은 성능을 나타내지만 영역 질의에 대해서 비교해야 하는 이동체 수가 늘어남에 따라 성능은 떨어지게 된다.

● 실험 인자 6 : R*-tree의 분기 계수(Branch Factor)

보고 즉시 색인 변경 정책을 사용할 경우 R*-tree의 분기 계수가 클수록 색인 변경 비용은 줄어드나, 한 노드 안에서 선형적으로 검색해야 할 객체수가 증가하기 때문에 영역 질의와 같은 공간 질의의 성능은 떨어진다

● 실험 인자 7 : 영역 질의의 영역 크기 비율

전체 실험 대상 데이터 집합의 영역에서 영역 질의의 크기의 비율로 나타낸다.

● 실험 1

이동체 수(인자1), 색인 변경 간격(인자2)의 값이 선형적으로 증가함에 따라 색인의 종류(인자3)와 색인 변경 정책(인자4)을 변경시켜가며, 색인의 변경 완료 때까지의 소요 시간을 측정한다.

실험을 통하여 색인 변경 정책 별로 이동체 수에 따른 색인 구축 간격과의 관계를 정형화 할 수 있다.

입력 인자	(인자 1, 인자2, 인자 3, 인자 4)
측정 값	색인의 변경 완료 때까지의 시간을 측정

● 실험 2

Grid File에서 이동체 수(인자1)와 버퍼 크기(인자5)가 선형적으로 증가할 경우 색인 변경 정책별 색인의 변경 완료 때까지의 시간을 측정한다.

실험을 통하여 Grid File에서 색인 변경 정책 별로 이동체 수와 버퍼 크기간의 관계를 정형화 할 수 있다.

입력 인자	(인자 1, 인자4, 인자 5)
측정 값	색인의 변경 완료 때까지의 시간을 측정

● 실험 3

R*-tree에서 이동체 수(인자1)와 분기 계수(인자6)가 선형적으로 증가할 경우 색인 정책별 색인 변경 완료 시간을 측정한다.

실험을 통하여 R*-tree에서 색인 변경 정책 별로 이동체 수와 분기 계수와의 관계를 정형화 할 수 있다.

입력 인자	(인자 1, 인자4, 인자 6)
측정 값	색인의 변경 완료 때까지의 시간을 측정

● 실험 4

영역 질의의 크기 비율(인자7)이 선형적으로 증가할 경우 색인의 종류(인자3) 별로 색인 변경 정책(인자4)을 변경시켜 가며 영역 질의의 응답 시간을 측정한다.

실험 4는 서로 다른 색인 변경 정책으로 구축된 색인에 대하여 검색 연산의 성능을 측정한다.

입력 인자	(인자3, 인자4, 인자 7)
측정 값	영역 질의의 응답 시간

4. 결론 및 향후 연구

이 논문은 이동체 공간 데이터베이스 시스템을 위한 이동체 표현 방법을 제안하고, 이를 바탕으로 한 공간 색인 변경 정책을 제시하였다. 제시한 변경 정책의 성능을 검증하기 위하여 성능 평가기를 제안하였으며, 현실 세계와 유사한 실험 환경을 제공하기 위해서 다양한 실험 인자를 정의하였다. 그 결과 색인 구축 간격이 비교적 작을 경우, Grid File에서는 위치 변경 보고 즉시 색인 변경 정책에서, R*-tree에서는 재 구축 방법에 각각 우수한 성능을 나타낼 것으로 예상된다

향후 보다 실제적인 실험을 위해서 실제 이동체의 움직임을 수집하여 실험해야 할 것이다.

5. 참고 문헌

- [1] O. Wolfson, S.Chamberlain, S. Dao, Liqin Jiang and Gisela Mendez, " Cost And Imprecision in Modeling the Poistion of Moving Object", Internatinal Conference Database Engineering 1998
- [2] O.Wolfson, B.Xu, S.Chamberlain and L.Jiang, " Moving Object Database : Issues and Solutions", In Proceedings of SSDBM Conference 2000. Pp. 123-132
- [3] Volker Gaede, " Multidimensional Access Method", ACM Computing Surveys, Vol.30, No.2, June 1998
- [4] Nievergelt J.,Hinterberger H. and Sevcik K. C., " The Grid File: An Adptable, Symmetric Multikey File Structure". 1984, ACM Transactions On Database System.. Vol.9, No1, pp38-71
- [5] Beckmann N., Kregel H.P., Schneider R, and Seeger B., "The R* Tree: An Efficient and Robust Access Method for Points and Rectangels", Proceedings of the ACM-SIGMOD Conference, Altantic City, pp 322-517
- [6] Jean-Marc Saglio, José Moreir, " : Oporto: A Realistic Scenario Generator for Moving Objects", Geominfomatica Volume 5, Number 1, March 2001, pp 71-93