

불확실 이동체의 질의 처리를 위한 불확실성 영역 기법

(Uncertainty Region Scheme for Query Processing of Uncertain Moving Objects)

반재훈[†] 홍봉희^{**} 김동현^{***}
 (Chaehoon Ban) (Bonghee Hong) (Donghyun Kim)

요약 위치기반서비스에서 이동체의 위치 데이터 수집 비용을 줄이기 위하여 위치 데이터를 주기적으로 수집한다. 주기적으로 수집된 위치 데이터는 보고 주기 사이의 위치 변화를 반영하지 못하기 때문에 시간에 대한 선형 함수를 이용하여 예측된 위치 데이터와 오차가 발생한다. 따라서 오차를 포함한 불확실한 미래위치데이터로 인하여 TPR 트리에서 현재위치질의 정확도가 떨어지는 문제점이 발생한다.

이 논문에서는 불확실한 위치 데이터에 대한 현재질의 처리를 위하여 선형 함수에 의해 예측된 위치 데이터에 오차분을 반영한 불확실성 영역을 정의하고 불확실성 영역을 설정하기 위하여 최근 예측 오차 가중치 기법과 칼만 필터 기법을 제시한다. 또한 TPR 트리를 기반으로 불확실성 영역을 반영한 질의 처리를 구현하고 성능 비교 평가를 수행한다. 성능 평가 결과에 따르면 기존의 선형함수 기반 질의처리 기법보다 불확실성 영역 기반 질의처리 기법이 최소 약 15% 이상의 정확도가 향상되는 장점을 가진다.

키워드 : 불확실 이동체, 이동체 데이터베이스, 지리정보시스템, 공간 색인

Abstract Positional data of moving objects can be regularly sampled in order to minimize the cost of data collection in LBS. Since position data which are regularly sampled cannot include the changes of position occurred between sampling periods, sampled position data differ from the data predicted by a time parameterized linear function. Uncertain position data caused by these differences make the accuracy of the range queries for present positions diminish in the TPR tree.

In this paper, we propose the uncertainty region to handle the range queries for uncertain position data. The uncertainty region is defined by the position data predicted by the time parameterized linear function and the estimated uncertainty error. We also present the weighted recent uncertainty error policy and the kalman filter policy to estimate the uncertainty error. For performance test, the query processor based by the uncertainty region is implemented in the TPR tree. The experiments show that the proposed query processing methods are more accurate than the existing method by 15%.

Key words : Uncertain Moving Objects, Moving Object Database, GIS, Spatial Indexing

1. 서론

무선 이동통신 기술의 발달로 휴대폰, PDA 등의 무선 이동기기가 보편화되면서 위치 기반 서비스(LBS, Location-Based Service)에 대한 요구가 증대되고 있

다. 특히 이동체 위치 파악 및 추적 서비스 분야의 응용 시스템 개발이 활발하게 진행되고 있으며 대표적인 예로 친구 찾기, 개인용 네비게이션(PNS, Personal Navigation System) 등이 있다. 응용 시스템에서 효과적으로 이동체에 대한 사용자 질의를 처리하기 위하여 이동체의 위치 정보를 관리하고 현재 위치 질의를 신속히 처리하는 방법이 필요하다.

이동체의 위치 데이터를 수집하기 위하여 크게 이동체의 이동 방향과 속도가 특정 임계값을 벗어나는 경우에 위치 데이터를 보고하는 방법과 주기적으로 위치 데이터를 보고하는 두 가지 방법이 있다. 임계값을 벗어날 때 마다 위치를 보고하는 전자의 방법은 도심의 차량과

* 이 논문은 교육인적자원부 지방연구중심대학육성사업(차세대물류IT기술연구사업단)의 지원에 의하여 연구되었음

[†] 정 회 원 : 경남정보대학 인터넷응용계열 교수
 chban@kit.ac.kr

^{**} 중신회원 : 부산대학교 컴퓨터공학과 교수
 bhhong@pusan.ac.kr

^{***} 정 회 원 : 동서대학교 디자인&IT 전문대학원 교수
 pusrover@dongseo.ac.kr

논문접수 : 2005년 9월 1일
 심사완료 : 2006년 2월 22일

같이 교통 상황에 따라서 움직임의 변화가 큰 이동체에 적용할 때 잦은 위치 보고로 이동체와 서버 사이의 통신비용이 증가하는 단점이 있다. 그러나 주기적으로 위치 데이터를 보고하는 후자의 방법은 이동체의 움직임의 변화에 관계없이 일정한 시간 간격마다 위치를 보고하기 때문에 위치 보고 횟수를 줄이고 위치 데이터 수집 비용을 감소하는 장점이 있다.

그러나 주기적으로 이동체의 위치 데이터를 수집하면 실제 위치 데이터와 색인 상에서 예측된 데이터간에 오차가 발생하는 문제점이 있다. 이동체의 현재 위치를 색인하는 TPR 트리(TPR-Tree)[1]는 이동체의 위치를 예측하기 위하여 위치 좌표와 속도를 매개변수로 하는 시간 기반의 선형 함수를 사용한다. 만약 주기적 데이터 수집 환경하에서 이동체를 색인하기 위하여 TPR 트리를 사용하면 시간에 대한 선형 함수를 이용하여 예측된 위치 데이터는 보고 주기 사이의 위치 변화를 반영하지 못하기 때문에 실제 수집되는 데이터와 오차가 발생하게 된다. 따라서 오차가 발생하는 예측 위치 데이터를 이용하여 TPR 트리에서 현재 위치 질의를 수행하면 질의의 정확도가 떨어지는 문제점이 발생한다.

이동체 위치의 불확실성을 반영한 기법으로 시간에 대한 선형 함수로 예측한 오차를 반영한 색인 연구로는 [2]가 있다. [2]에서는 기 설정된 샘플링 오차와 측정 오차 값을 이용하여 색인과 질의 영역에서의 불확실성을 처리하였다. 그러나 이 연구에서는 구체적인 불확실성 영역을 설정하기 위하여 미리 이동체가 가지는 오차의 크기를 알고 있어야 하기 때문에 이동체의 이동 특성을 반영하지 못하는 문제점이 있다.

이 논문에서는 주기적인 위치 보고 환경에서 불확실한 위치 데이터에 대한 현재질의를 처리하기 위하여 선형 함수에 의해 예측된 위치 데이터에 오차분을 반영한 불확실성 영역을 정의한다. 그리고 이동체의 불확실성 영역을 설정하기 위하여 과거의 이동체의 위치 예측 오차를 이용한 최근 예측 오차 가중치 기법과 칼만 필터 기반 기법을 제시한다. 또한 불확실성 영역을 기반으로 확장된 TPR 트리를 이용하여 성능 비교 평가를 수행한다. 성능 평가 결과에 따르면 기존의 선형함수 기반 질의처리 기법보다 불확실성 영역 기반 질의처리 기법이 최소 약 15% 이상의 정확도가 향상되는 장점을 가진다.

이 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고 3장에서는 문제 정의를 기술한다. 4장에서는 이동체 위치의 불확실성 처리에 대해서 설명하고 5장에서는 불확실성 영역 설정 방법에 대해서 설명한다. 그리고 6장에서는 불확실성 기반 이동체 현재 위치 색인에 대한 설명을 하고 7장에서 실험을 통한 성능 평가 결과를 분석하고 8장에서는 결론 및 향후 연구를 기술한다.

2. 관련 연구

이동체의 현재 위치 질의를 위한 연구로는 이동체의 현재 및 가까운 미래 위치 표현을 위한 데이터 모델을 제시한 연구[3-6]와 이동체의 현재 및 가까운 미래 위치 질의에 대해 색인에 관한 연구[7-10]가 진행되어 왔다.

시간에 따라 연속적으로 변하는 이동체의 위치를 모두 데이터베이스에 저장할 수가 없기 때문에 이동체 데이터베이스는 이산적인 위치 데이터를 저장하게 된다. 이산적인 위치 데이터를 기반으로 이동체의 현재 위치를 추정하기 위해서 위치를 시간에 대한 선형 함수로 표현한 기법이 주로 [3-6]에서 연구되었다. [3-6]은 선형 함수에 의한 데이터 모델에서 갱신에 대한 비용모델과 위치 갱신 정책을 제안하였다.

선형함수 기반의 데이터 모델링을 이용해서 이동체의 현재 및 미래 위치에 대한 검색을 위한 색인 기법으로 TPR 트리[1]가 있다. TPR 트리는 R* 트리 구조를 기반으로 이동체의 현재 위치를 시간매개변수최소경계사각형(TPBR, Time Parameterized Bounding Rectangle)을 이용하여 색인한다. TPBR은 노드의 모든 엔트리들을 포함하는 시간 함수 기반의 경계 사각형이다. TPBR의 각 축은 내부 이동체들의 이동 방향에 따라 가장 빠른 속도 값으로 증가하기 때문에 이동체의 위치가 선형함수로 예측한 위치에서 임계 영역내에 있음을 보장한다. TPR 트리는 주기적으로 보고된 이동체의 위치와 선형 함수에 의한 위치 예측 간의 오차를 반영하지 못하는 문제점이 있다. TPR 트리 기반의 연구로는 TPR* 트리[8], STAR 트리[9], R-exp 트리[10]가 있다. 이 색인들은 TPR 트리에 비해서 색인의 검색 성능을 향상시켰지만 동일한 데이터 모델을 사용하기 때문에 이동체의 위치 오차를 반영하지 않고 있다.

이동체의 위치 불확실성에 관한 연구로는 크게 과거 궤적 선분에 대한 연구[6,1]와 현재 위치에 대한 연구[2]가 있다. 이동체의 과거 위치 데이터는 곡선의 형태를 가진 연속적인 움직임을 이산적인 선분의 형태로 나누어서 표현하기 때문에 두 끝점을 연결한 선분은 보고되지 않은 끝점 사이의 위치에서 불확실성이 발생하게 된다. [6,1]에서는 이동체의 과거 위치 데이터의 선분화에 따른 불확실성을 해결하기 위하여 해당 선분의 양 끝점 및 이웃 선분의 데이터를 이용한 선형 보간법을 제안하였다. 그러나 한쪽 끝점의 위치를 알지 못하는 현재 위치에 대한 불확실성에는 적용할 수 없는 단점이 있다. [2]는 현재위치 데이터에 대하여 시간의 선형 함수로 예측한 오차를 반영한 연구로서 이동체의 불확실성 요인을 측정 오차와 샘플링 오차로 구분해서 반영한다. 측정 오차는 미리 설정이 가능한 오차이지만, 샘플링 오차는

기존의 선형 함수에 미리 설정한 오차를 반영하여 이동체의 불확실성을 처리한다. 그러나 불확실성을 계산하기 위하여 미리 오차율을 설정하며 위치를 보고할 때마다 변하는 위치 예측의 오차를 반영하지 못하는 단점이 있다.

예측된 값의 오차를 보정하는 대표적인 기법으로 칼만필터[12,13]가 있다. 칼만필터는 각 단계별로 값을 예측하고 예측된 값을 동적으로 보정하는 동적 예측모델로서 예측단계에서 상태 방정식을 이용하여 현 상태의 상태 값을 결정한다. 현 상태의 값을 결정할 때 이전 상태에서부터 칼만필터가 적용된 시스템의 오차를 합한다. 이를 위하여 기본적으로 두 개의 방정식인 출력 방정식과 상태방정식으로 구성되며 상태방정식을 통하여 나온 상태벡터를 이용하여 현재 시점의 출력벡터를 구할 수 있다. 칼만필터를 이용한 위치오차예측 방법은 5.3절에서 자세히 기술한다.

3. 문제 정의

3.1 이동체의 위치보고 방법

일반 도로를 따라 움직이는 이동체는 시간과 위치에 따른 도로의 상황 변화에 따라 움직임의 변화가 크게 발생한다. 그러므로 속도와 방향이 바뀔 때 마다 위치를 보고하는 방법은 많은 통신 비용을 발생시키고, 위치 보고 횟수를 예측할 수 없으므로 통신 비용을 예측하기가 어렵다.

예를 들면 그림 1과 같이 이동체가 화살표를 따라서 이동한다고 할 때, 속도와 방향이 바뀔 때 마다 위치보고를 하는 경우에는 그림 1의 (a)와 같이 x표시가 되어 있는 부분에서 모두 위치 보고를 해야 하지만, 주기적으로 위치보고를 하는 경우에는 그림 1의 (b)와 같이 설정된 시간마다 보고를 함으로써 보고 횟수를 줄일 수

있다. 특히, 도로의 상황이 정체와 소통을 반복하는 경우와 경사가 많은 지역일수록 그 차이 커진다고 할 수 있다. 또한 주기적으로 이동체의 위치를 보고하는 방법은 위치 보고 횟수를 정확히 알 수 있으므로 통신 비용의 예측이 가능하다는 장점이 있다.

3.2 위치 오차의 경우

이동체는 그림 2와 같이 3가지 경우의 오차가 발생한다. 경우 1은 위치 측정 시 오차이고 경우 2는 속도 측정 시 발생하는 오차이다. 그리고 경우 3은 위치 보고 주기 사이에 발생하는 속도와 방향의 변화에 따른 이동체 예측 위치의 오차이다. 경우 1과 경우 2는 GPS와 같은 위치 및 속도 데이터 수집 장비에 의한 측정오차이다. 그러므로 이 논문에서는 경우 1과 경우 2에 대한 오차는 고려하지 않는다. 그러나 경우 3은 이동체가 보고하는 주기 사이의 위치 변화를 반영하지 못해서 발생하는 오차이다.

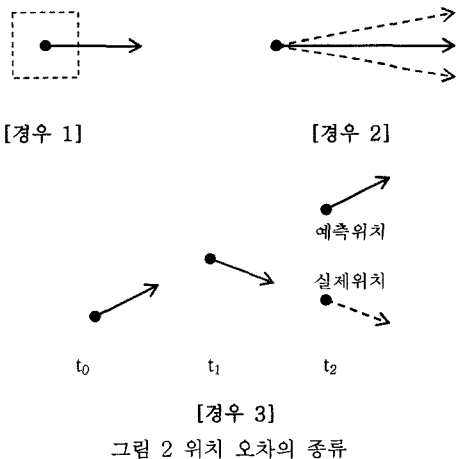
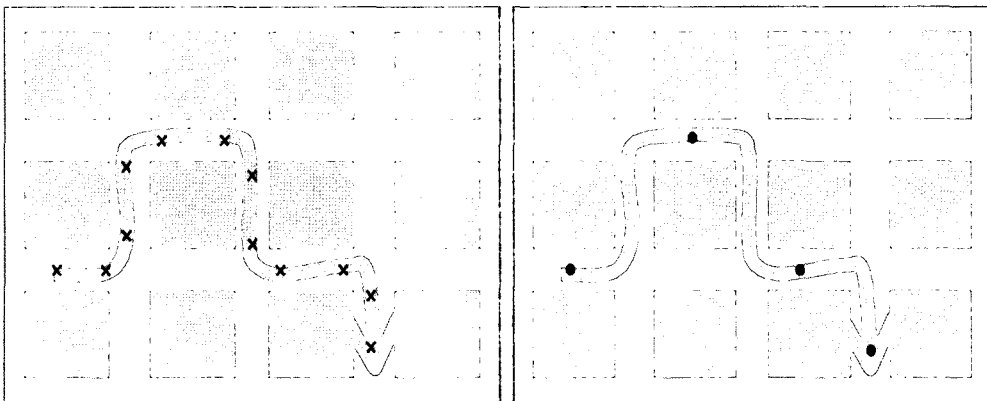


그림 2 위치 오차의 종류



(a) 속도와 방향의 변화에 따른 위치보고

(b) 주기적인 위치보고

그림 1 위치보고 방법

3.3 위치 오차에 따른 문제점

그림 3은 주기적으로 위치 보고하는 이동체에서 보고 주기 사이에 발생하는 오차의 예를 나타낸다. 이동체의 위치 보고를 한 뒤에 이동체의 속도와 방향이 바뀌는 구간에서 위치 보고가 이루어지지 않으면 이동체의 실제 위치는 B지점이지만 선형 함수에 의해서 예측된 위치는 A지점이므로 위치 예측의 오차가 크게 발생하게 된다. 그러므로 영역1에 대하여 영역 질의를 하게 되면 실제로 존재하지 않는 이동체를 찾게 되는 문제점이 발생한다. 이를 오류적중(false hit)라고 한다. 만약 영역2에 질의를 하면 비록 이동체가 영역2에 존재하지만 검색 결과에 포함되지 않는다. 이를 오류누락(false miss)라 한다.

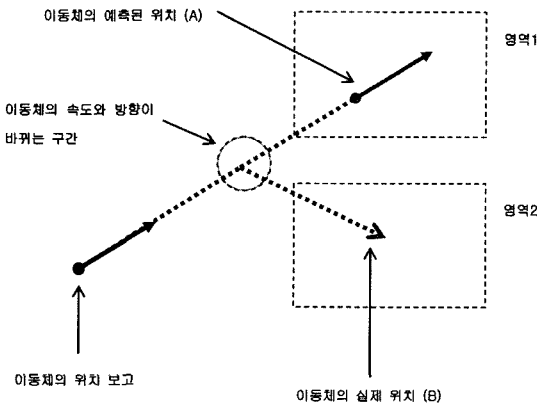


그림 3 주기적인 위치보고 시 발생하는 오차

전체 이동체의 집합을 U라 하고 시공간 영역 S에 대한 영역 검색 질의를 처리하기 위한 색인의 정확한 후보결과집합을 R 그리고 오차가 발생한 후보결과집합을 P라 하자. 이 때 오류적중은 오차가 발생한 후보결과집합에 속하지만 정확한 후보결과집합에 속하지 않은 객체이며 오류적중집합은 P-R이다. 오류누락은 후보결과집합들에 포함되어야 하지만 오차로 인하여 누락된 객체이며 오류누락집합은 R-P이다. 따라서 오류적중비율(false hit ratio)과 오류누락비율(false miss ratio)은 식 (1)과 같이 계산된다.

$$false\ hit\ ratio = \frac{n(P-R)}{n(P)} \times 100$$

$$false\ miss\ ratio = \frac{n(R-P)}{n(R)} \times 100 \quad (1)$$

주기적으로 이동체의 위치를 보고할 때 TPR 트리와 같이 선형함수로 이동체의 현재 위치를 예측하는 색인은 예측 위치의 계산 시 이동체의 불확실성을 반영하지 못한다. 또한 단말 노드의 시간매개변수 최소경계사각형

이 특정 시간에 대한 임계영역만을 포함한 매우 작은 영역으로 계산되기 때문에 결과집합에 속해야 하는 이동체를 검색하지 못하는 오류누락비율이 증가하는 문제점이 발생한다. 이 논문에서는 TPR 트리의 시간매개변수 최소경계사각형을 계산할 때 불확실성 영역을 정의하고 설정함으로써 오류적중비율을 최소한으로 유지하면서 오류누락비율을 개선한다.

4. 이동체 위치의 불확실성 처리

4.1 불확실성

이동체는 보고 시점의 공간 좌표 x_i, y_i 와 공간상의 각축 방향으로의 속도인 V_{x_i}, V_{y_i} 그리고 보고 시점의 시간 t_i 를 보고한다. 따라서 이동체의 i 번째 보고된 위치 정보 $f_i = (x_i, y_i, V_{x_i}, V_{y_i}, t_i)$ 으로 표현된다. 이동체의 예측 위치 P_i 를 시간에 대한 선형함수로 예측한 위치라 할 때 t_i 시간의 이동체의 예측 위치 $P_i = (P_{x_i}, P_{y_i})$ 로 표현된다. 이 때 t_i 시간의 P_{x_i}, P_{y_i} 는 식 (2)와 같이 정의된다.

$$P_{x_i} = x_{i-1} + V_{x_{i-1}} * (t_i - t_{i-1})$$

$$P_{y_i} = y_{i-1} + V_{y_{i-1}} * (t_i - t_{i-1}) \quad (2)$$

이동체의 i 번째 실제 위치를 $R_i = (R_{x_i}, R_{y_i})$ 라 할 때 위치 예측의 오차는 그림 4와 같이 이동체가 실제 존재하는 위치 R_i 와 예측한 위치 P_i 의 차이이다.

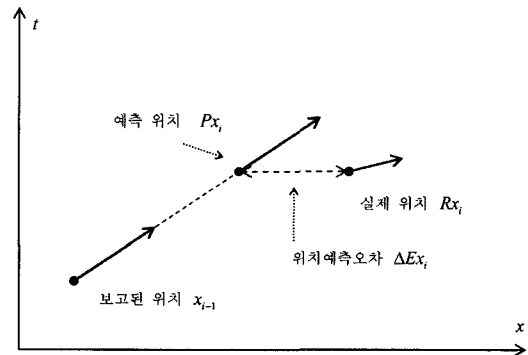


그림 4 위치 예측의 오차

i 번째 위치 예측의 오차 ΔE_i 를 $(\Delta E_{x_i}, \Delta E_{y_i})$ 라 할 때 ΔE_i 는 식 (3)과 같이 정의된다.

$$\Delta E_{x_i} = R_{x_i} - P_{x_i}$$

$$\Delta E_{y_i} = R_{y_i} - P_{y_i} \quad (3)$$

주기적인 위치보고 환경에서 이동체의 위치는 선형함수에 의해서 예측할 수 없기 때문에 이동체의 불확실성은 $\Delta E_x \neq 0$ 또는 $\Delta E_y \neq 0$ 인 상태이다. 따라서 이동체 불확실성은 보고 주기 사이에 발생하는 실제 존재하는 위치와 선형함수에 의해 예측한 위치의 오차에 대한 속도 벡터로 정의할 수 있으며 i 번째 불확실성을 $U_i = (Ux_i, Uy_i)$ 으로 나타낸다. 이 때 오차는 공간상의 4 방향으로 발생하기 때문에 실제 i 번째 이동체의 불확실성 $U_i = (Ux_i^{\rightarrow}, Ux_i^{\leftarrow}, Uy_i^{\rightarrow}, Uy_i^{\leftarrow})$ 로 정의된다.

정의 1. 불확실성(uncertainty): 이동체 위치 예측의 불확실성으로 시간에 대한 선형함수로 예측한 위치와 실제 보고된 이동체 위치의 오차이다. x 축 방향의 오차 속도벡터를 각각 $Ux_i^{\rightarrow}, Ux_i^{\leftarrow}$ 라 하고 y 축 방향의 오차 속도벡터를 각각 $Uy_i^{\rightarrow}, Uy_i^{\leftarrow}$ 라 할 때 이동체의 i 번째 불확실성 U_i 는 다음과 같이 정의된다.

$$-U_i = (Ux_i, Uy_i) = (Ux_i^{\rightarrow}, Ux_i^{\leftarrow}, Uy_i^{\rightarrow}, Uy_i^{\leftarrow})$$

4.2 불확실성 영역

불확실성 영역(UR, Uncertainty Region)은 이동체 위치 예측의 불확실성을 반영한 영역으로서, 시간에 대한 선형함수로 예측한 위치에 이동체의 위치 예측의 오차를 반영한 영역이다. 예를 들어 그림 5와 같이 $i-1$ 시간에 보고된 이동체의 실제위치 R_{i-1} 과 선형함수 $f(x)$ 를 이용하여 i 시간의 예측위치 P_i 를 계산할 수 있다. 이 때 그림 5의 점선으로 된 영역과 같이 P_i 에서 과거의 위치 예측 오차를 계산하여 각 축 방향으로 확장한 영역이 i 시간의 이동체에 대한 불확실성 영역이다. 즉 공간의 각 축으로 이동체 위치 예측 오차를 현재 이동체가 이동할 수 있는 영역으로 설정함으로써 이동체 위치의 불확실성을 설정한다.

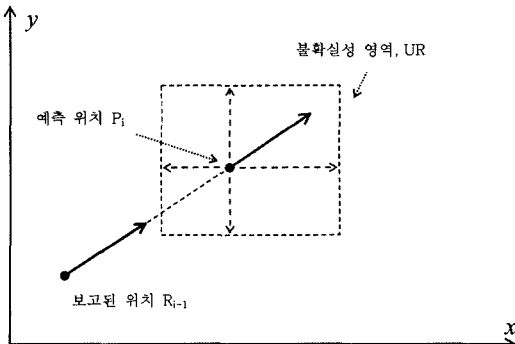


그림 5 불확실성 영역의 예

그림 6과 같이 $i-1$ 시간에 보고된 이동체의 위치 R_{i-1} 과 선형함수 $f(x)$ 를 이용하여 i 시간의 위치 P_i 를 예측한다. 그러나 이 예측된 위치는 실제 위치 R_i 와 다르다. 불확실성 영역은 이 예측된 위치 P_i 에서 확장한 영역으로 실제 위치 R_i 를 포함하기 위해서 확장한다. 그림 6에서 i 시간에 질의 영역 S에 대한 질의가 들어올 경우, 선형 함수로 위치를 예측하는 경우에는 실제 위치 R_i 가 질의 영역에 포함되더라도 예측 위치가 질의 영역에 포함되지 않으므로 검색할 수가 없다. 그러나 이동체의 위치 예측을 불확실성 영역을 사용할 경우에는 질의 영역 S와 불확실성 영역 UR_i 가 겹치기 때문에 이동체를 검색할 수 있다.

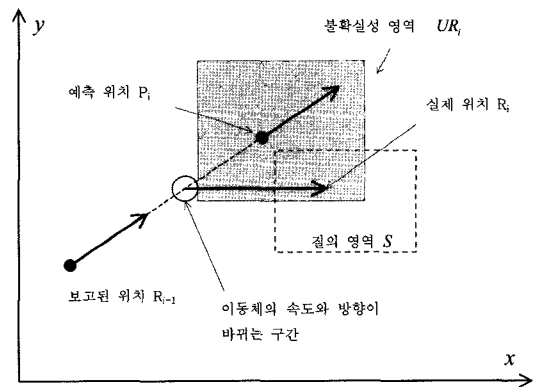


그림 6 불확실성 영역의 효과

불확실성 영역은 $UR_i = (UR_{x_i}^{\rightarrow}, UR_{x_i}^{\leftarrow}, UR_{y_i}^{\rightarrow}, UR_{y_i}^{\leftarrow})$ 로 나타내고 다음 식 (4)와 같이 정의한다.

정의 2. 불확실성 영역(uncertainty region, UR): 이동체 위치 예측의 불확실성을 표현한 영역으로 시간 매개 변수의 선형함수를 이용하여 예측한 위치에 예측의 오차를 추가하여 확장한 영역이다.

$$UR_i = (UR_{x_i}^{\rightarrow}, UR_{x_i}^{\leftarrow}, UR_{y_i}^{\rightarrow}, UR_{y_i}^{\leftarrow})$$

$$UR_{x_i}^{\rightarrow} = P_{x_i} - U_{x_i}^{\leftarrow} = x_{i-1} + V_{x_{i-1}} * (t_i - t_{i-1}) - U_{x_i}^{\leftarrow} * (t_i - t_{i-1})$$

$$= x_{i-1} + (V_{x_{i-1}} - U_{x_i}^{\leftarrow}) * (t_i - t_{i-1})$$

$$UR_{x_i}^{\leftarrow} = P_{x_i} - U_{x_i}^{\rightarrow} = x_{i-1} + V_{x_{i-1}} * (t_i - t_{i-1}) - U_{x_i}^{\rightarrow} * (t_i - t_{i-1})$$

$$= x_{i-1} + (V_{x_{i-1}} - U_{x_i}^{\rightarrow}) * (t_i - t_{i-1})$$

$$UR_{y_i}^{\rightarrow} = P_{y_i} - U_{y_i}^{\leftarrow} = y_{i-1} + V_{y_{i-1}} * (t_i - t_{i-1}) - U_{y_i}^{\leftarrow} * (t_i - t_{i-1})$$

$$= y_{i-1} + (V_{y_{i-1}} - U_{y_i}^{\leftarrow}) * (t_i - t_{i-1})$$

$$UR_{y_i}^x = P_{y_i} - U_{y_i}^x = y_{i-1} + V_{y_{i-1}} * (t_i - t_{i-1}) - U_{y_i}^x * (t_i - t_{i-1})$$

$$= y_{i-1} + (V_{y_{i-1}} - U_{y_i}^x) * (t_i - t_{i-1}) \quad (4)$$

5. 불확실성 영역 설정 정책

이 장에서는 불확실성 영역의 설정하기 위한 최근 예측 오차 가중치 정책과 칼만 필터를 적용한 정책을 기술한다.

5.1 불확실성 영역의 설정

이동체의 위치 예측은 예측 시간이 커질수록 오차가 커진다. 그러므로 불확실성 영역은 시간이 지남에 따라서 커지도록 설정되어야 한다. 불확실성 영역을 시간에 관계없이 일정한 크기로 설정할 경우에는 다음의 그림 7과 같은 문제점이 발생한다. 그림 7에서와 같이 이동체 A는 t_0 시간에 접선으로 된 화살표 방향인 $f(t)$ 방향으로 B 또는 C로 이동한다. 이때 이동체 A에 대한 불확실성 영역을 UR로 미리 설정할 경우에 t_1 시간에는 이동체 A에 대한 불확실성을 반영할 수 있지만, t_0 시간에는 위치 오차에 비해서 불확실성 영역이 너무 크게 설정되고, t_2 시간에는 위치 오차에 비해서 불확실성 영역의 크기가 너무 작게 설정된다. 그러므로 불확실성 영역은 위치 보고 시간과 예측 시간 사이의 시간의 크기를 반영하여야 한다.

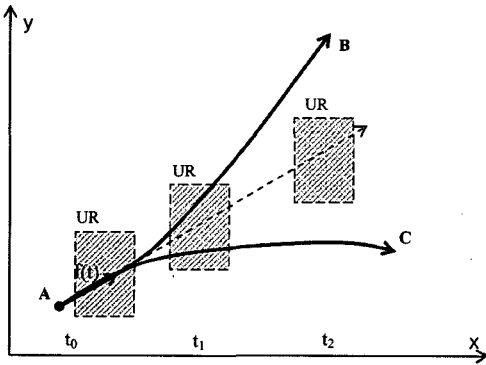


그림 7 고정 크기 방식의 문제점

그림 8에서와 같이 이동체 A가 t_0 시간에 접선으로 된 화살표 방향인 $f(t)$ 방향으로 이동할 때, 이동체 A의 방향전환에 따른 $f(t)$ 와의 위치 오차는 시간이 지남에 따라서 커지게 된다. 즉 이동체 A가 B방향으로 가는 경우와 C방향으로 가는 경우 모두 t_1 시간보다 $f(t)$ 와의 차이가 커지게 된다. 시간이 지남에 따라서 이동체의 위치예측의 오차가 커지는 것을 반영하기 위해서 불확실성 영역은 시간에 따라서 선형적으로 증가하도록 한다.

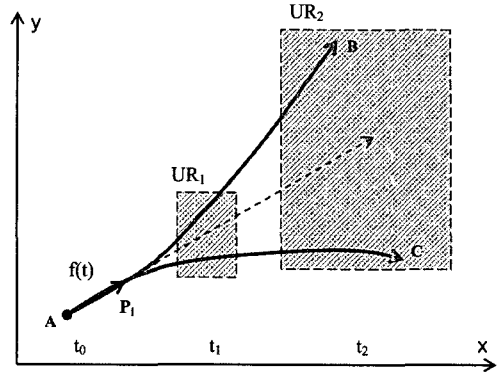


그림 8 시간에 따른 영역크기 변화의 효과

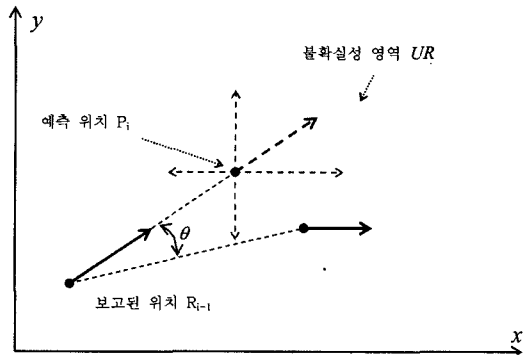


그림 9 방향 오차

그림 8과 같이 t_0 시간부터 $f(t)$ 방향으로 이동하는 이동체에 대한 불확실성 영역은 시간이 t_1, t_2 로 증가할 때마다 불확실성 영역의 크기 역시 UR_1, UR_2 로 확대되어야 한다.

이동체의 위치예측의 오차는 속도의 변화뿐만 아니라 방향의 변화도 같이 발생한다. 그러므로 속도와 방향의 변화를 모두 반영할 수 있는 불확실성 영역 설정이 필요하다. 속도의 변화를 반영하기 위해서는 공간상의 두 방향의 속도오차만 반영하면 된다. 이는 불확실성 영역의 속도값이 벡터의 형태로 사용되기 때문이다. 그러나 방향의 변화를 반영하기 위해서는 공간상의 네 방향의 속도성분 모두를 반영해야 한다. 그림 9에서와 같이 이동체의 예측위치와 실제위치 사이에는 θ 의 각도로 오차가 발생한다. 이와 같이 각도의 변화가 발생할 경우에는 불확실성 영역이 전 방향으로 같은 크기로 확장되어야 할 필요가 없다. 불확실성 영역은 각도의 변화가 작은 부분은 크게 확장될 필요가 없이 각도의 변화가 많은 부분으로 더 크게 확장한다.

5.2 최근 예측 오차 가중치 정책

이동체의 이동방향 및 속도는 시간상으로 최근의 방향 및 속도에 영향을 많이 받는 특성을 가지고 있다. 최

근 예측 오차 가중치 정책은 이 특성을 이용하여 최근의 위치 예측 오차에 더 많은 가중치를 두고 오차를 계산하는 정책이다. 예를 들어 예측 오차는 이동체가 이동하는 지역에 따라서 동일한 지역에서는 유사한 오차 크기로 발생한다. 이동체가 A지역 → B지역 → C지역 → D지역으로 이동하고 있고 각 지역에서 발생하는 오차는 다르다고 가정하자. 이때 이동체가 C지역에 있다면 위치 예측의 오차는 C지역에서의 발생하는 위치예측의 오차와 유사한 크기로 발생한다. 즉, A지역이 고속도로이고 C지역이 2차선 도로일 경우에, 이동체가 A지역에 있을 동안에는 위치 예측의 오차가 거의 발생하지 않는다. 이는 고속도로와 같은 도로에서는 속도와 방향의 변화량이 적기 때문에 이동체의 불확실성도 줄어들기 때문이다. 그러나 C지역과 같은 2차선 도로는 잦은 속도와 방향의 변화를 발생시키므로 불확실성의 크기도 커진다. 최근 예측오차 가중치는 이러한 특성을 반영하고자 최근에 발생한 예측오차에 큰 가중치를 두어서 이동체의 불확실성을 반영하는 방법이다.

최근 예측 오차 가중치 정책에서는 최근 예측 오차에 가중치를 두기 위한 기준 매개변수인 factor 변수를 이용한다. factor 변수는 불확실성 영역 계산시 최근에 발생한 오차의 반영 범위를 결정하며 0부터 1사이의 값을 가지게 된다.

$i-1$ 번째 위치 보고를 한 t_{i-1} 시간의 이동체의 위치 예측의 오차를 ΔE_{i-1} 라고 할 때 factor를 이용한 t_i 시간의 예측 오차 U_i 는 식 (5)와 같이 계산된다.

$$U_i = \Delta E_{i-1} * factor + U_{i-1} * (1 - factor) \quad (5)$$

이 때 i 번째 위치보고의 예측 오차는 식 (6)과 같이 계산된다.

$$\begin{aligned} U_2 &= \Delta E_1 * factor \\ U_3 &= \Delta E_2 * factor + U_2 * (1 - factor) \\ &\dots \\ U_i &= (\Delta E_{i-1} + \Delta E_{i-2} * (1 - factor)^1 + \Delta E_{i-3} * (1 - factor)^2 + \\ &\quad \dots + \Delta E_1 * (1 - factor)^{i-2}) * factor \\ &= \left(\sum_{n=1}^{i-1} (\Delta E_n * (1 - factor)^{i-n-1}) \right) * factor \end{aligned} \quad (6)$$

식 (6)에 따르면 최근 예측 오차 가중치 정책에서 가장 최근의 예측 오차는 factor의 크기만큼 반영이 되고 과거 n 번째 예측 오차는 $factor * (1 - factor)^{i-n-1}$ 만큼 반영이 된다. 따라서 factor의 값이 1에 가까울수록 최근의 예측 오차를 많이 반영하고 0에 가까울수록 과거에 발생한 예측 오차를 많이 반영한다.

5.3 칼만 필터 기반 기법

칼만 필터는 시간의 흐름에 따라 값을 추정하고, 추정 한 값의 오차를 보정하는 동적 예측 모델이다. 칼만 필터 기반 기법은 먼저 상태 방정식을 이용하여 이전 상태에서부터 시스템의 내적 오차를 합하여 현 상태의 상태 벡터를 결정한다. 그리고 새로운 이동체의 위치보고가 입력된 위치 예측의 오차, 동적 회귀계수 역할을 하는 상태 벡터 그리고 출력오차의 합으로 관측치(현 상태의 위치 예측의 오차를 추정한 값)를 계산한다.

칼만 필터의 기본적인 모델구성은 다음과 같다.

$$Y_t = F_t \theta_t + v_t \quad (\text{출력식})$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (\text{상태식})$$

Y_t 는 관측 값으로 이루어진 출력벡터이고 θ_t 는 출력 벡터인 F_t 를 통해서 Y_t 와 대응되는 값으로 관측되지 않는 값인 동적계수벡터 즉, 상태벡터이다. G_t 는 t 시점에 상태벡터인 θ_t 의 변화를 나타내기 위한 전이행렬이며 v_t 는 출력 오차(measurement noise)이다. 마지막으로 w_t 는 상태오차(process noise)를 나타낸다.

칼만 필터는 시간의 흐름에 따라 주어진 초기정보와 상태 방정식으로 동적계수벡터 θ_t 값들을 생성하고 출력 방정식에서 동적계수벡터를 이용하여 출력 벡터 Y_t 를 생성하는 구조를 지닌 모형이다. 그리고 t 시점의 관측치 Y_t , 출력 벡터 F_t , 전이행렬 G_t , 오차 분산들이 결정된 상태에서 $t+1$ 시점의 최적 동적계수값들을 추정해 가는 동적 순환 과정이 반복적으로 이루어진다.

칼만 필터를 기반으로 불확실성 영역을 계산할 때 F_t 는 알고 있는 정보들을 나타내는 값이고 Y_{t+1} 의 값이 불확실성 영역이 된다. 불확실성 영역 $UR_t = (URx_t^{\sim}, URx_t^{\sim}, URY_t^{\sim}, URY_t^{\sim})$ 로 정의내기 때문에 x, y 축으로 최대, 최소 방향의 경계 값을 모두 계산하여야 한다. 그러므로 공간 상의 4방향에 대한 불확실성을 모두 입력 값으로 칼만 필터를 통해서 예측해야 한다. 이를 위하여 칼만 필터의 입력 값으로 i 시간에 발생한 이동체의 불확실성인 $(Ux_i^{\sim}, Ux_i^{\sim}, Uy_i^{\sim}, Uy_i^{\sim})$ 을 각각 입력하고 결과값으로 불확실성 영역 설정에 필요한 i 시간까지 누적된 이동체의 불확실성 정보를 출력한다.

6. 실험

6.1 실험 데이터

실험을 위하여 이동체 생성을 위한 대표적인 도구인 GSTD(Generate Spatio-Temporal Data)[11]를 이용하

여 데이터 집합을 생성하였다. 각 데이터 집합은 랜덤 분포와 가우시안 분포의 데이터 집합이며 하나의 데이터 집합은 1,000개의 이동체가 1,000회 위치 보고를하도록 설정하였다.

주기적인 위치 보고환경에서의 실험을 위해서 각 데이터 집합의 타임스텝 주기를 달리하면서 실험을 하였다. 실험 시 타임스텝 주기는 5, 10, 15, 20, 25주기로 설정하였으며 질의는 공간 축으로 5%, 10%, 20% 영역 질의를 사용하여 주기의 변화에 의한 성능 변화를 측정하였다. 또한 GSTD에 의해서 생성되는 데이터는 시공간상의 점으로 표현되지만 성능평가에 적용하기 위해서는 위치 좌표와 속도 벡터가 삽입되므로 동일 이동체 id를 가진 시간으로 순서화된 두 점을 이용해서 속도 벡터를 나타내어 성능 평가에 사용하였다. GSTD를 이용하여 생성한 데이터 집합의 분포도는 그림 10과 같다.

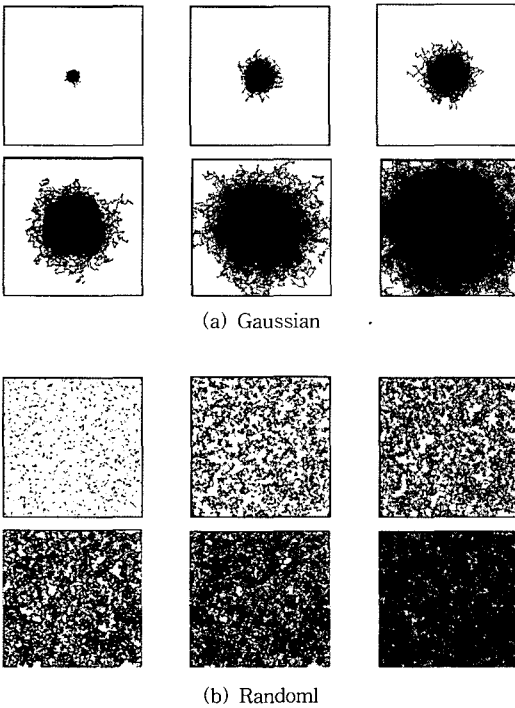


그림 10 데이터 집합 분포도

6.2 질의 정확도 비교

각 데이터 집합에 대한 질의의 정확도를 계산하기 위하여 선형 함수 기반 기법[2], 칼만 필터 기반 기법, 최근예측오차 가중치 기반 기법에 대하여 오류적중율(false hit ratio)와 오류누락율(false miss ratio)을 측정하였다.

그림 11은 랜덤 분포의 데이터 집합에 대한 오류적중율을 보여준다. 그림 11에서 보듯이 5% 영역 질의에 있어서는 3가지 기법이 모두 거의 유사한 적중율을 보이지만 질의 영역이 10%와 20%로 커질수록 칼만 필터와 최근 예측오차 가중치 기법을 이용했을 때 약 4~5% 정도의 근소한 성능 향상을 보이고 있다.

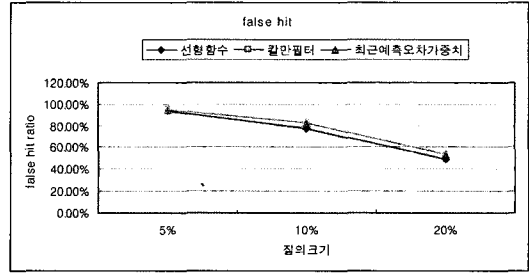


그림 11 Random1 데이터 집합에 대한 오류적중율

그림 12는 랜덤 분포의 데이터 집합에 대한 오류누락율을 보여준다. 오류누락율은 5% 질의에서는 선형 함수 기반 기법이 90% 이상이며 칼만 필터와 최근 예측오차 가중치 기반 기법은 약 76%~78%선이다. 질의 영역이 커질수록 검색에 대한 정확도는 모두 높아져서 10% 영역 질의에 대해서 선형 함수 기반 기법이 약 79% 그리고 칼만 필터와 최근 예측오차 가중치는 약 58%의 오류누락율이 발생한다. 또한 20% 영역질의에 대해서는 선형 함수 기반 기법이 약 50% 그리고 칼만 필터와 최근 예측오차 가중치 기반 기법이 약 22%~24%의 오류누락율이 발생한다. 이는 질의 영역이 커질수록 이동체의 위치예측의 오차 크기가 질의 영역내에 존재할 확률이 커지므로 오류누락율이 작아지기 때문이다. 그림 12에서 보듯이 칼만 필터와 최근 예측오차 가중치 기반 기법이 선형 함수 기반 기법에 비해서 약 15%에서 27%의 성능 향상을 보이고 있다.

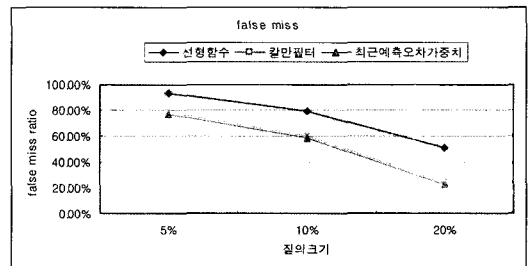


그림 12 Random1 데이터 집합에 대한 오류누락율

6.3 보고 주기에 따른 성능 비교

주기적인 위치보고 환경에 이동체의 불확실성은 주기

가 길수록 커지는 특성을 가지고 있다. 보고 주기에 따른 정확도를 측정하기 위해서 보고 주기의 크기를 변화하면서 실험을 하였다. 보고 주기는 GSTD 타임스탬프 2,5,10,15,20,25 주기로 설정해서 실험을 하였다.

그림 13과 그림 14는 가우시안 분포의 데이터 집합에 대하여 보고 주기의 변화에 따른 선형 함수, 칼만 필터 그리고 최근예측오차 가중치 기반 기법을 이용할 때의 성능비교 실험 결과를 보여준다. 그림 13에서 보듯이 모든 3가지 기법에 대하여 오류적중율이 2타임스탬프에서 50%정도이지만 10타임스탬프를 넘어가면 90%이상으로 높아지는 단점을 보인다. 이유는 세가지 기법 모두 질의 영역은 일정하지만 보고 주기가 길수록 불확실성 영역이 커지기 때문에 높은 오류적중률을 보이는 것으로 분석된다. 이 현상은 그림 11에서 보았듯이 데이터 분포와 무관하게 보여지기 때문에 향후 개선되어야 할 부분이다. 그림 14에서 보듯이 오류누락율 역시 10타임스탬프부터 성능이 크게 하락한다. 그러나 칼만 필터와 최근예측오차 가중치 기법은 선형 함수 기법에 비해서 2타임스탬프 주기에서 약 20%, 10타임스탬프까지는 약 40% 그리고 10타임스탬프가 넘어가면 약 30% 정도 오류누락율이 작아진다.

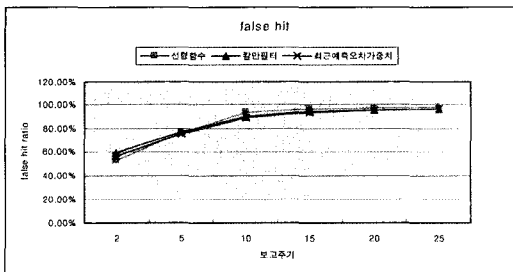


그림 13 보고 주기에 따른 오류적중율

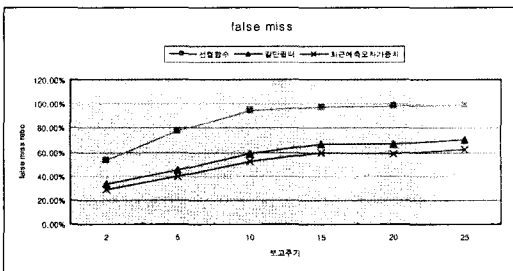


그림 14 보고 주기에 따른 오류누락율

8. 결론 및 향후 연구

주기적으로 이동체의 위치데이터를 보고할 때 위치 보고 횟수를 줄이고 위치 데이터 수집 비용을 감소하는

장점이 있다. 그러나 이동체의 현재 위치를 색인하기 위하여 TPR 트리 TPR 트리를 사용하면 수집된 위치 데이터는 보고 주기 사이의 위치 변화를 반영하지 못하기 때문에 시간에 대한 선형 함수를 이용하여 예측된 위치 데이터와 오차가 발생하게 된다. 따라서 현재위치 질의를 수행하면 질의의 정확도가 떨어지는 문제점이 발생한다

이 논문에서는 주기적으로 위치를 보고하는 환경에서 이동체 위치예측의 오차로 인한 영역 검색의 정확도가 저하되는 문제점을 개선하기 위해서 이동체의 불확실성을 반영한 불확실성 영역을 정의하였다. 그리고 불확실성 영역을 계산하기 위하여 최근 예측 오차에 가중치를 두는 기법과 칼만 필터를 이용하는 기법을 이용하였다. 최근 예측오차 가중치 방법과 칼만 필터를 적용한 기법은 과거에 발생한 예측오차 정보를 이용해서 적절한 오차의 크기를 추정하며 속도와 방향에 대한 오차를 모두 반영한다. 성능 평가 결과에 따르면 기존의 선형함수 기반 기법보다 불확실성 영역을 이용한 기법이 최소 약 15% 이상의 정확도가 향상되는 장점을 가졌다.

향후 연구로 좀 더 실제 이동체와 유사한 움직임은 가지는 위치데이터에 대한 추가적인 실험과 검색 결과의 정확도에 대한 비용모델에 대한 연구가 필요하다. 그리고 오류적중률에 대한 실험 결과에서 보듯이 검색의 정확도를 더욱 향상시키기 위한 방법과 색인의 성능 향상에 대한 추가적인 연구가 필요하다.

참고 문헌

- [1] S. Saltenis, C. S. Jensen, S.T. Leutenegger, and M. A. Lopez, "Indexing the Positions of Continuously Moving Objects," Int'l Conf. on Management of Data and Symposium on Principles Databases and Systems, pp.331-342, 2000.
- [2] J.H.Hosbond, S.Saltenis and R.Ørtoft, "Indexing Uncertainty of Continuously Moving Objects," Int'l Conf. on Database and Expert Applications, pp.911-915, 2003.
- [3] A.P.Sistla, O.Wolfson, S.Chamberlain, and S.Dao, "Querying the Uncertain Position of Moving Objects," Lecture Notes in Computer Science, Vol.1399, pp.310-337, 1997.
- [4] O. Wolfson, B.Xu, S. Chanmberlain, and L. Jiang, "Moving objects database: issues and solutions," Int'l Conf. on Scientific and Statistical Database Management, pp.111-122, 1998.
- [5] D.Pfoser and C.S. Jensen, "Capturing the Uncertainty of Moving-Object Representations," Int'l Symposium on Large Spatial Databases, pp.111-132, 1999.
- [6] D.Pfoser and Nectaria Tryfona "Capturing Fuzziness and Uncertainty of Spatiotemporal Objects,"

- Advances in Databases and Information Systems, pp.112-126, 2001.
- [7] N. Beckmann and H. P. Kriegel, "The *R*-tree*: An Efficient and Robust Access Method for Points and Rectangles," Int'l Conf. on Management of Data and Symposium on Principles Databases and Systems, pp. 332-331, 1990.
- [8] Y.Tao, D.Papadias and J.Sun, "The *TPR*-tree*: An Optimized Spatio-Temporal Access Method for Predictive Queries," Int. Conf. on Very Large Data Bases, pp.790-801, 2003.
- [9] C.M Procopiuc, K. Agarwal and Sarel Har-Peled, "STAR-Tree: An Efficient Self-Adjusting Index for Moving Objects," Workshop on Algorithm Engineering & Experiments, pp.178-193, 2002.
- [10] Simonas Saltenis and C.S. Jensen, "Indexing of Moving Objects for Location-Based Services," Int'l Conf. on Data Engineering, pp.463-472, 2002.
- [11] Tomas Brinkhoff, "Generating Network-Based Moving Objects," Int'l Conf. on Scientific and Statistical Database Management, pp.253-255, 2000.
- [12] Kalman,R.E "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME-Journal of Basic Engineering, pp.35-45, 1960.
- [13] West, M and Harrison, J, *Bayesian Forecasting and Dynamic Models*, 2nd Ed., Springer, 1997.

반 재 훈

정보과학회논문지 : 데이터베이스
제 33 권 제 1 호 참조

홍 봉 희

정보과학회논문지 : 데이터베이스
제 33 권 제 1 호 참조



김 동 현

1995년 부산대학교 컴퓨터공학과(학사)
1997년 부산대학교 컴퓨터공학과(석사)
2003년 부산대학교 컴퓨터공학과(박사)
2004년~현재 동서대학교 디자인&IT 전문대학원 조교수. 관심분야는 지리정보시스템, 모바일 GIS, 모바일 트랜잭션, 이

동체 색인