

시공간 유사성을 이용한 도로 네트워크 상의 유사한 궤적 검색

황 정 래[†] · 강 혜 영^{**} · 이 기 준^{***}

요 약

이동 객체들의 행동을 분석하기 위해서는 그 이동 객체 궤적들 간의 유사성 측정을 정의하는 것이 필요하다. 비록 유클리디언 공간 상에서 이동 객체들의 유사한 궤적들을 검색하는 몇몇 연구가 있었지만, 도로 네트워크 공간 상에서 이동 객체들의 궤적에 대한 유사성 연구는 거의 관심을 끌지 못 했다. 실제 응용에서, 대부분의 이동 객체들은 유클리디언 공간보다 도로 네트워크 공간 상에 존재한다. 그러나 궤적들 간의 유사성 측정에서, 기존의 방법들은 유클리디언 거리를 기반으로 하고 있으며, 오직 공간적 유사성만을 고려하고 있다. 본 논문에서, 우리는 도로 네트워크 상에서 POI와 TOI를 기반으로 유사성 측정을 정의한다. 이러한 정의를 바탕으로, 우리는 궤적들 간의 시공간 유사성을 사용하여 유사한 궤적을 검색하는 방법들을 제시한다. 그리고 유사한 궤적들에 대하여 클러스터링을 수행한 결과를 보인다. 실험 결과는 각 방법에 의해 검색된 유사한 궤적들과 그 검색된 궤적 결과들이 각 방법 간에 얼마나 일치하는지를 보인다.

키워드 : 궤적, 도로 네트워크 공간, 궤적들 간의 유사성

Similar Trajectory Retrieval on Road Networks using Spatio-Temporal Similarity

Jung-Rae Hwang[†] · Hye-Young Kang^{**} · Ki-Joune Li^{***}

ABSTRACT

In order to analyze the behavior of moving objects, a measure for determining the similarity of trajectories needs to be defined. Although research has been conducted that retrieved similar trajectories of moving objects in Euclidean space, very little research has been conducted on moving objects in the space defined by road networks. In terms of real applications, most moving objects are located in road network space rather than in Euclidean space. In similarity measure between trajectories, however, previous methods were based on Euclidean distance and only considered spatial similarity. In this paper, we define similarity measure based on POI and TOI in road network space. With this definition, we present methods to retrieve similar trajectories using spatio-temporal similarity between trajectories. We show clustering results for similar trajectories. Experimental results show that similar trajectories searched by each method and consistency rate between each method for the searched trajectories.

Key Words : Trajectories, Road Network Space, Similarity between Trajectories

1. 서 론

모바일 컴퓨팅의 확산에 따라 이동 객체들을 효과적으로 다루기 위한 연구가 중요한 이슈가 되고 있다. 이러한 이동 객체들의 움직임은 (x, y, t) 공간에서 라인 세그먼트의 집합으로 된 궤적으로 표현되어진다[1]. 이동 객체 궤적은 많은 정보를 포함하고 있기 때문에, 몇몇 응용 분야에 대해서 궤적을 분석하는 것은 흥미있는 일이다. 이동 객체 궤적들을

분석하는 가장 중요한 필요성의 하나는 이동 객체들의 유사한 궤적과 그들의 클러스터링을 검색하는 것이다. 예를 들면, “주어진 질의 궤적과 유사한 궤적을 가지는 모든 이동 객체들을 찾아라”와 같은 질의는 궤적들 상에서 전형적인 질의의 한 형태이다. 그러나 우리는 아무런 의미없는 위치와 시간에 관심을 두지 않는다. 우리는 도로 상의 수많은 관심이 있거나 중요한 지점들을 고려한다. 심지어, 우리는 도로 상에 움직이는 이동 객체들의 시간대에도 관심을 가진다. 따라서, 도로 네트워크 상의 이동 객체 궤적들을 관심있는 지점(Points Of Interest)과 관심있는 시간대(Times Of Interest)로 표현하는 것이 의미가 있을지도 모른다. 예를 들어, 두 이동 객체가 도로 네트워크 상의 동일한 POIs와 동

* 본 논문은 산업자원부 한국산업기술평가원 지정 한국항공대학교 부설 인터넷정보검색 연구센터에 의하여 지원되었습니다.

[†] 준 회 원 : 부산대학교 지형정보학과 박사수료

^{**} 준 회 원 : 부산대학교 컴퓨터공학과 박사수료

^{***} 정 회 원 : 부산대학교 컴퓨터공학과 교수

논문접수 : 2005년 9월 14일, 심사완료 : 2006년 5월 8일

일한 TOI를 통과하였다면, 우리는 이 두 이동 객체 궤적이 유사하다라고 간주할 수 있다.

비록 이동 객체들의 유사한 궤적을 검색하는 몇몇 연구가 유클리디언 공간 상에서 이루어졌지만, 도로 네트워크 상의 이동 객체들에 대해서는 거의 주목을 받지 못했다. 그러나 대부분의 실제 응용에 대해서, 우리는 유클리디언 공간보다 도로 네트워크 공간에서의 이동 객체들에 관심이 더 많다. 따라서 이동 객체들의 행동을 분석하기 위해서, 우리는 이동 객체들의 궤적들 간의 유사성 측정을 정의하는 것이 필요하다. 이러한 유사성 측정은 유사한 궤적들을 검색할 뿐만 아니라 유사한 궤적들을 클러스터링함으로써 도로 정보, 궤적 정보 그리고 차량 정보 등의 연계를 통하여 타겟 마케팅과 같은 실제 응용에 사용되어질 수 있다.

도로 네트워크 공간의 특징들에 기인하여, 도로 네트워크 공간 상에서 유사한 궤적들을 검색하는 방법들은 다음과 같은 결점을 가지는 기존의 방법들과는 다르다[2, 3]. 첫째, 기존의 방법들은 유클리디언 공간을 가정으로 한다. 그러나 유클리디언 공간은 도로를 따라 정의되어지는 거리를 가진 도로 네트워크 공간 상에서는 타당하지 않다. 궤적들의 유사성 측정은 거리의 정의에 의해 결정되기 때문에 유클리디언 공간에 대해 정의된 유사성 측정은 도로 네트워크 공간에 적합하지 않다. 따라서, 유클리디언 공간을 기반으로 한 유사한 궤적들을 검색하는 방법들은 본 논문의 목적에 적절하지 않다.

둘째, 기존의 방법들은 궤적의 시공간 특성들을 완전히 활용하지 않고 대부분의 방법들이 오직 공간적 유사성만 고려한다. 예를 들면, 두 궤적이 서로 다른 시간대에 동일한 지역을 통과할 때, 시공간적으로 볼 때 두 궤적은 유사하지 않다. 심지어, 두 이동 객체 궤적이 서로 다른 방향의 움직임을 가지더라도, 이 궤적들은 공간적 유사성만을 고려하는 기존의 방법에 의해 유사하다라고 간주되어진다.

이러한 기존 방법들의 결점을 극복하기 위하여, 우리의 연구는 두 가지 요건에 의해 동기가 주어진다. 첫째, 우리의 방법은 도로 네트워크 상의 공간적 특성과 이동 객체들의 특성을 바탕으로 한다. 둘째, 우리는 공간적 유사성뿐만 아니라 시공간적 유사성도 함께 고려한다. 이러한 요건들을 기반으로, 우리는 도로 네트워크 상의 이동 객체들의 유사한 궤적들에 대한 검색 방법을 제시한다. 우리의 방법은 시공간적 특성을 기반으로 하고, 도로 네트워크 상의 공간적 특성을 반영한다. 그리고 본 논문에서 제시한 우리의 방법은 도로 네트워크 상의 궤적들과 운전자 속성들의 통합과 함께 지리 마케팅이나 보험업과 같은 실제 응용에 매우 관심있고 유용하게 사용될 것이다

본 논문의 구성은 다음과 같다. 2장에서, 우리는 관련 연구와 기존 방법들의 문제점을 소개하고 도로 네트워크 상의 이동 객체들의 궤적에 대한 유사성의 특징에 대하여 살펴본다. 3장에서, 우리는 도로 네트워크 상의 유사한 궤적들을 검색하는 방법을 제시한다. 4장에서, 우리는 검색된 유사한 궤적들에 대하여 수행된 클러스터링 결과를 보인다. 5장에

서, 우리는 이동 객체 궤적 데이터를 사용하여 실험한 결과를 보인다. 마지막으로, 우리는 6장에서 결론을 맺고 향후 연구를 제시한다.

2. 관련 연구 및 동기

본 장에서, 우리는 도로 네트워크 상의 이동 객체들과 관련된 연구들을 소개한다. 그리고, 우리는 도로 네트워크 상의 이동 객체들의 궤적의 특성을 조사함으로써 기존 방법들의 문제점을 토론하고 본 논문의 동기 부여를 제시한다.

2.1 관련연구

이동 객체들의 유사한 궤적들을 검색하는 것은 두 가지 쟁점 - 1) 이동 객체 궤적의 표현 방법 2) 궤적들 간의 유사성의 정의 - 과 밀접한 연관을 가진다.

첫번째 연구 쟁점과 관련하여, 이동 객체 궤적에 대한 표현이 많은 연구에서 이루어져 오고 있다[4, 5]. 특히, 궤적들에 대한 표현 모델이 [6]에서 Markovian과 non-Markovian 확률 모델 기반으로 제시되었다. 이것은 궤적들로부터 유용한 정보를 추출할 때 효과적이다. 다른 관심 있는 모델이 다중 입상의 지리공간적 생명선을 고려하여 [7]에서 제시되었다. 이러한 방법들은 유클리디언 공간 상의 이동 객체들을 다룬다. 그러나 실생활에서, 자동차나 기차와 같은 대부분의 이동 객체들은 유클리디언 공간보다 도로 네트워크 공간에서 주로 찾아볼 수 있다. 따라서 도로 네트워크 상의 이동 객체를 효과적으로 표현하기 위한 몇몇 연구가 이루어졌다. 도로 네트워크 상의 이동 객체들의 표현과 질의를 위한 간결한 모델이 [5]에서 명확히 정의되었으며, 도로 네트워크를 따라 이동하는 제한된 움직임의 환경에서 이동 객체 궤적들에 대하여 표현하고 추론하는 접근이 [8, 9]에서 제시되었다.

두번째 연구 쟁점과 관련하여, 유사한 궤적들에 대한 검색 방법과 관련된 중요한 연구들이 [2]와 [10]에서 찾아볼 수 있다. [2]와 [10]에서 정의된 두 경로 A 와 B 간의 유사성 함수 $S1$ 과 $S2$ 는 다음과 같다.

$$S1(\delta, A, B) = \frac{SigmoidSim_{\delta}(A, B)}{min(n, m)}$$

여기서, n 과 m 은 각각의 경로 A 와 B 에 대한 점의 개수이다. $SigmoidSim$ 는 시간 확장을 허락한다. 따라서, 만약 시간대들이 가까이 있으면, 서로 다른 시간대에서 공간에 가깝게 있는 객체들은 일치될 수 있다. $S1$ 은 평행 움직임을 찾아내기 위한 유사성 함수이다.

$$S2(\delta, A, B) = \max_{f_{c,d} \in F} S1(\delta, A, f_{c,d}(B))$$

여기서, F 는 이차원에서 변형의 집합이라 하자. 그 때, 만

약 $f_{cd}(A) = ((a_{x,1} + c, a_{y,1} + d), \dots, (a_{x,n} + c, a_{y,n} + d))$ 이면 함수 f_{cd} 는 F 에 속한다.

위 정의들을 기반으로 가장 긴 공통된 서브시퀀스 모델 (Longest Common Subsequence Model)을 이용하여 데이터베이스에서 주어진 질의 궤적과 가장 유사한 궤적을 찾는 방법이 [10]에서 제시되었다. 그러나 이 방법은 도로 네트워크 상의 이동 객체들의 유사한 궤적을 찾는 데 두 가지 문제점을 가지고 있다. 먼저, 이 방법은 시간적 또는 시공간적 측면을 고려하지 않는다. 예를 들면, 두 이동 객체 궤적이 서로 다른 시간대에 동일한 지역을 통과하더라도 [10]에서 제시된 방법은 그 두 궤적이 유사하다고 간주한다. 둘째, [10]의 방법이 유클리디언 공간을 기반으로 하기 때문에 도로 네트워크 상의 유사한 궤적을 검색하는데 적용하기가 힘들다.

모양을 기반으로 한 궤적의 유사성 측정이 [3]에서 정의되었다. 동일한 시간적 길이 ($L_T(\lambda) = L_T(\lambda')$)와 자연수 m 을 가진 두 경로 λ 와 λ' 에 대하여, 두 경로 λ 와 λ' 간의 시공간 유사성 $D_{TS}(\lambda, \lambda')$ 은 [3]에서 다음과 같이 정의되었다.

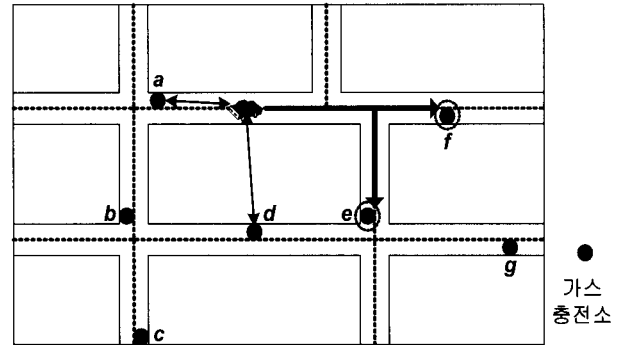
$$D_{TS}(\lambda, \lambda') = \frac{1}{m+1} \sqrt{\sum_{i=0}^m D(X_{\lambda_n}(i), X_{\lambda'_n}(i))^2}, \text{ where } \Delta t = \frac{L_T(\lambda)}{m} = \frac{L_T(\lambda')}{m}$$

여기서, $\lambda_{\Delta t}$ 는 λ 를 동일한 간격 Δt 로 나눔으로써 생성되어진다.

[3]에서 정의된 유사성의 장점은 [2]와는 다르게 시공간적 측면이 고려되었다는 것이다. 그러나 이 방법 역시 유클리디언 공간을 가정으로 하기 때문에 도로 네트워크 공간에 적용하기가 어렵다. 그리고 [11]과 [12]에서도 유사한 방법이 제시되었으나 [2]에서 제시된 방법과 같이 유클리디언 거리나 타임 와핑 거리의 문제점을 가짐으로써 도로 네트워크 상에 적용하기가 힘들다. 그외 도로 네트워크 상의 이동 객체들에 대한 근접 이웃 검색 방법이 [16-18, 20]에서 소개되었으며, 이동 객체 궤적들에 대한 다양한 질의 방법들이 [19]에서 제시되었다.

2.2 도로 네트워크 상의 이동 객체 궤적의 유사성

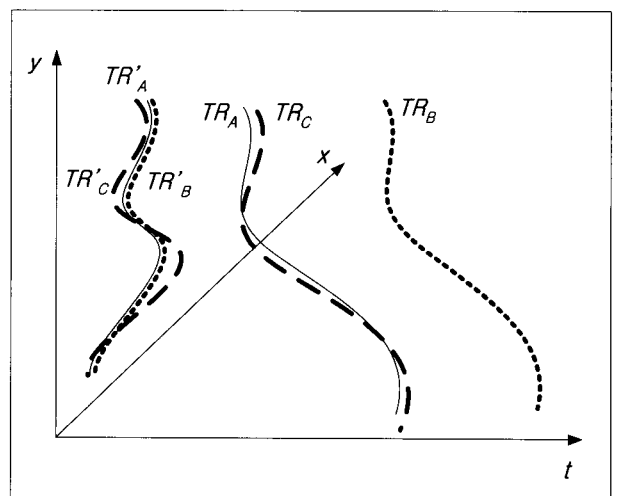
실생활에서 대부분의 이동 객체들은 유클리디언 공간보다 도로 네트워크 공간 상을 움직인다. 유클리디언 공간과 도로 네트워크 공간 간에는 몇 가지 차이점이 있다. 첫째, (그림 1)은 유클리디언 공간과 도로 네트워크 공간 간의 서로 다른 거리의 정의를 설명하고 있다. (그림 1)에서와 같이 오른쪽으로 움직이고 있는 차량에서 가장 가까운 두 개의 가스 충전소를 찾는다 가정하자. 이때, 만약 우리가 유클리디언 거리로 찾았다면 a 와 d 가 찾아진다. 하지만, 도로 네트워크를 따라 찾았다면 e 와 f 가 찾아진다. 여기서, 우리는 유클리디언 거리로 찾은 a 와 c 는 도로 네트워크를 고려할 때 주어진 차량으로부터 의미가 없는 가스 충전소라는 것을 알 수 있다.



(그림 1) 유클리디언 공간과 도로 네트워크 공간 간의 거리의 차이

둘째, 다른 좌표 시스템이 도로 네트워크 공간에 사용된다. (x, y, t) 좌표 시스템이 유클리디언 공간에 가장 인기 있는 좌표 시스템의 하나인 반면에, (Sid, d, t) 은 도로 네트워크 공간에 더 효율적인 좌표 시스템의 하나이다. 여기서, Sid 는 도로 섹터의 식별자이고, d 는 도로 섹터의 출발점으로부터의 거리를 의미한다. 예를 들면, 도로 네트워크 상의 질의는 유클리디언 공간에서와 같이 어떤 한 지역을 지정하는 것 보다 비교적 도로 섹터 ID를 지정함으로써 주어진다. 이것은 유클리디언 좌표 시스템보다 도로 네트워크 좌표 시스템을 사용함으로써 도로 네트워크 상의 두 지점 간의 거리를 계산하기가 더 쉽다는 것을 의미한다. 마지막으로, 도로 네트워크 공간은 도로 세그먼트 간의 연결성을 묘사하기 위하여 추가적 데이터를 요구한다. 이러한 차이점들은 도로 네트워크 공간에서 궤적들을 분석할 때 주의깊게 조사되고 고려되어야 한다.

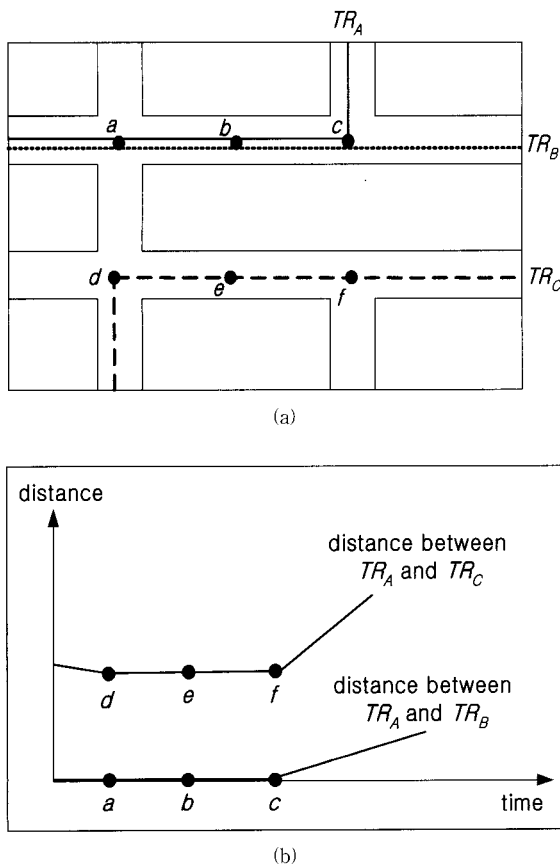
우리는 도로 네트워크 상의 궤적들의 특징들에 대하여 살펴보도록 한다. (그림 2)는 (x, y, t) 공간 상의 궤적들의 예를 보여주고 있다. 여기서, t 는 시간축을 나타내고 (x, y) 공간은 투영된 공간이다. (그림 2)에서 TR_A, TR_B 그리고 TR_C 는 (x, y, t) 공간에서의 궤적들인 반면에 TR'_A, TR'_B 그리고 TR'_C 는 (x, y) 평면으로 투영된 궤



(그림 2) 시공간 공간에서의 궤적과 그것들의 투영된 궤적

적들이다. TR_A 와 TR_B 는 정확히 동일한 지점들을 통과하지만 서로 다른 시간 간격을 가지고 통과한다. TR_A 와 TR_C 는 서로 약간 떨어져 통과하지만 거의 비슷한 시간 간격을 가지고 움직인다. 우리는 TR_A , TR_B 그리고 TR_C 를 (x, y) 평면으로 투영하였을 때, TR_A 와 TR_B 는 (x, y) 평면 상에 동일한 궤적으로 투영되어진다. 반면에 TR_A 와 TR_C 의 투영된 궤적들은 서로 다른 위치에 놓여진다. 이것은 유사성 측정에서 시공간적 측면이 고려되어진다면, TR_C 는 오직 투영된 궤적들 간의 유사성을 고려하는 [2]와 [11]에서 제시된 유사성 측정과 다르게 TR_B 보다 TR_A 와 더 유사한 궤적으로 간주되어진다는 것을 의미한다.

도로 네트워크 공간 상의 이동 객체들의 거리는 흥미있는 특징을 가진다. 두 이동 객체의 궤적 TR_A 와 TR_B 가 동일한 시간에 도로 상의 동일한 지점 a, b 그리고 c 를 통과하고 그것들은 (그림 3)의 (a)에 묘사되었듯이 서로 다른 도로로 들어간다고 가정하자. 이때, (그림 3)의 (b)에서 보여지듯이 그 궤적들 간의 거리는 도로 네트워크에 따라 c 지점을 지난 후에 급격히 증가한다. 대부분의 경우, 서로 다른 두 도로 섹터 상에 있는 두 이동 객체 간에는 비교적 큰 거리가 생기고, 이에 따라 (그림 3)(b)에서 보듯이 유사성 검색에 대한 거리의 한계를 초과한다. 이러한 관찰은 만약 두 이동 객체가 서로 다른 도로 섹터 상에 있다면, 두 이동 객체 간의 거리를 계산하는 것은 의미가 없을지도 모른다는 것을 함축하고 있다.



(그림 3) 도로 네트워크 상의 두 이동 객체 간의 거리의 변화

3. 도로 네트워크 상의 유사한 궤적들의 검색

도로 네트워크 상의 유사한 궤적들을 검색하기 위하여, 우리는 다음과 같은 방법들의 하나를 활용할 수 있다.

- 방법 1 : 공간적 유사성을 기반으로 궤적들을 여과하고 시간적 거리를 기반으로 유사한 궤적들을 정제하는 방법
- 방법 2 : 시간적 유사성을 기반으로 궤적들을 여과하고 공간적 거리를 기반으로 유사한 궤적들을 정제하는 방법
- 방법 3 : 궤적들 간의 시공간 거리를 기반으로 유사한 궤적들을 검색하는 방법

우리는 이러한 방법들의 각각에 대하여 자세히 살펴보도록 한다.

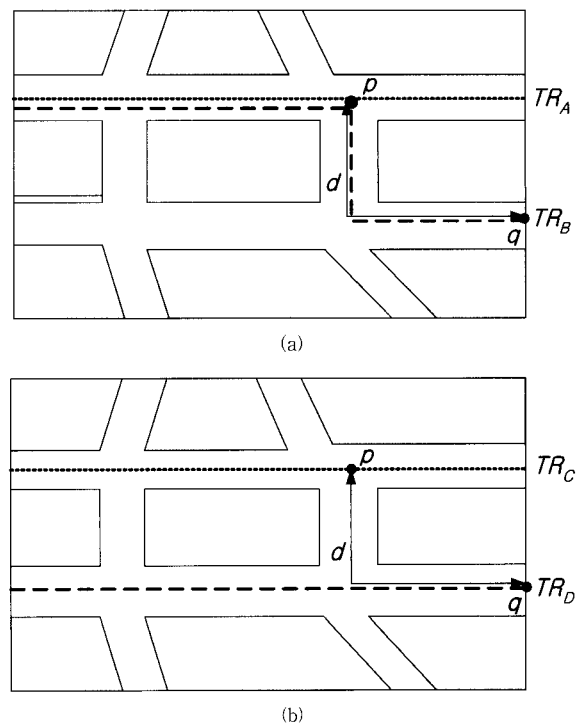
3.1 공간적 필터링과 시간적 거리를 기반으로 유사한 궤적들을 검색

이 방법에 대해, 우리는 먼저 곡선으로 표현되어지는 궤적들 간의 공간적 거리의 정의가 필요하다. 널리 사용되는 곡선 l 과 m 간의 거리의 정의인 하우스도르프 거리 $dist_H(l, m)$ 는 다음과 같이 정의되어진다.

$$dist_H(l, m) = \max_{a \in l} (\min_{b \in m} dist(a, b))$$

여기서, $dist(a, b)$ 는 두 지점 간의 거리이다.

하우스도르프 거리의 관심있는 특징은 도로 네트워크 상의 궤적들을 찾아내는 것이다. (그림 4)의 (a)에서, TR_A 와



(그림 4) 도로 네트워크 상의 하우스도르프 거리

TR_B 간의 하우스도르프 거리 $dist_H(TR_A, TR_B)$ 는 d 로 점 p 와 q 의 쌍으로 결정되어진다. 그리고 우리는 TR_C 와 TR_D 간의 하우스도르프 거리 $dist_H(TR_C, TR_D)$ 도 d 라는 것을 안다. 즉, 응용 형태에 따라 조금씩 다를 수 있지만 TR_A 와 TR_B 간의 하우스도르프 거리는 TR_C 와 TR_D 간의 거리와 동일하다. 이것은 우리가 도로 네트워크 공간에서 공간적 거리의 측정으로 하우스도르프 거리를 사용할 수 없다는 것을 의미한다.

하우스도르프 거리를 대신하여, 우리는 POI(Points of Interest)를 기반으로 하여 궤적들 간의 공간적 유사성을 결정하는 실용적인 방법을 제시한다. POI는 도로 네트워크 상에 존재하는 관심이 있거나 중요한 장소로 여기는 지점을 의미한다. 예를 들어, 공항, 학교 그리고 교차로 등과 같은 장소가 POI가 될 수 있다. 만약 두 궤적이 동일한 POI들을 통과한다면, 그것들은 다음의 정의에 의해 유사성이 고려되어진다.

[정의 1] 도로 네트워크 공간 상의 궤적들 간의 공간적 유사성

주어진 도로 네트워크 상의 POI들의 집합을 P 라 하자. 이때, 두 궤적 TR_A 와 TR_B 간의 공간적 유사성은 다음과 같이 정의되어진다.

$$Sim (TR_A, TR_B, P) = \begin{cases} 1, & \text{만약 } \forall p \in P, p \text{가 } TR_A \text{와 } TR_B \\ & \text{상에 있는 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

이 방법을 적용하기 위하여, 공간적 유사성 외에 시간적 유사성이 필요하다. 시간적 유사성은 시간적 거리의 정반대로 정의되어진다. 이 시간적 거리는 하나의 POI가 주어질 때 동일한 POI에서 통과 시간의 차이로 다음과 같이 정의되어진다.

[정의 2] 하나의 POI에 대한 궤적들 간의 시간적 거리

$p \in P$ 그리고 P 는 POI의 집합이라고 하자. 이때, 두 궤적 TR_A 와 TR_B 간의 시간적 거리는 다음과 같다.

$$dist_T (TR_A, TR_B, p) = |t (TR_A, p) - t (TR_B, p)|$$

만약 TR_A 와 TR_B 가 p 를 통과하지 않으면, 시간적 거리는 무한대로 간주되어진다.

만약 우리가 i 번째 POI에서의 통과 시간을 $t(TR, p_i)$ 로 간주하면, 각 궤적 TR 은 k 차원 공간에 있는 점 $t(TR) = (t(TR, p_1), t(TR, p_2), \dots, t(TR, p_k))$ 으로 나타내어진다. 여기서, k 는 POI들의 수이다. 이때, POI들의 집합에 대한 두 궤적 간의 시간적 거리는 이 k 차원 공간의 L_p 거리로 다음과 같다.

알고리즘 1. 도로 네트워크 상의 유사한 궤적들의 검색 : 방법

```

Input.    input trajectories  $TR_{IN}$ , threshold  $\epsilon$ ,
          query trajectory  $tr_Q$ , POI set  $P$ 
Output.   similar trajectories  $TR_{OUT}$ 
Begin
    // initialize
     $TR_{Candidate} \leftarrow \emptyset$ 
     $TR_{OUT} \leftarrow \emptyset$ 
    // filtering step
    For each  $tr \in TR_{IN}$ 
        If  $\forall p \in P, p$  is on  $tr$ 
            then  $TR_{Candidate} \leftarrow TR_{Candidate} \cup \{tr\}$ 
    // refinement step
    For each  $tr \in TR_{Candidate}$ ,
        If  $dist_T (tr_Q, tr, P) < \epsilon$ 
    then  $TR_{OUT} \leftarrow TR_{OUT} \cup \{tr\}$ 
    return  $TR_{OUT}$ 
End
    
```

[정의 3] POI들의 집합에 대한 궤적들 간의 시간적 거리

POI의 집합을 P 라 하고 두 궤적을 TR_A 와 TR_B 라 하자. 이때, TR_A 와 TR_B 간의 시간적 거리는 다음과 같이 정의되어진다.

$$dist_T (TR_A, TR_B, P) = L_p (TR_A, TR_B, P) = (\sum_{i=1}^k |(p_i (TR_A) - (p_i (TR_B))|^p)^{1/p}$$

알고리즘 1은 본 절에서 설명된 도로 네트워크 상의 유사한 궤적들을 검색하는 과정을 요약하고 있다.

이 알고리즘은 공간적 유사성을 기반으로 한 여과 단계와 시간적 거리를 기반으로 유사한 궤적들을 검색하는 정제 단계의 두 단계로 구성되어 있다.

3.2 시간적 필터링과 공간적 거리를 기반으로 유사한 궤적들을 검색

도로 네트워크 상의 궤적들 간의 유사성 검색에서 오직 시간적 유사성만을 고려한다면 그것은 아무런 의미가 없을 지도 모른다. 하지만 시간적 유사성을 여과한 후, 여과된 궤적 데이터들로부터 공간적 거리를 기반으로 유사한 궤적들을 검색한다면 그것은 의미가 있을 수 있다. 본 절에서, 도로 네트워크 상의 유사한 궤적 검색을 위하여, 우리는 시간적 유사성을 기반으로 궤적들을 여과하고 공간적 거리를 기반으로 유사한 궤적들을 정제하는 방법을 살펴본다. 만약 두 이동 객체 궤적이 동일한 POI를 통과한다면, 그것들은 다음의 정의에 의하여 유사성이 고려되어진다.

[정의 4] 도로 네트워크 상의 궤적들 간의 시간적 유사성

주어진 도로 네트워크 상의 POI들의 집합을 T 라고 하자. 이때, 두 궤적 TR_A 와 TR_B 간의 시간적 유사성은 다음과 같이 정의되어진다.

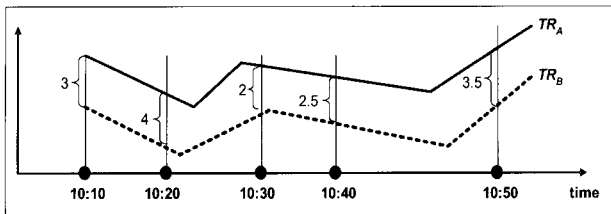
$$Sim (TR_A, TR_B, t) = \begin{cases} 1, & \text{만약 } t \in [t_s(TR_A), t_e(TR_A)] \ \& \ t \in [t_s(TR_B), t_e(TR_B)] \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

시간적 유사성을 기반으로 여과된 궤적들을 대상으로, 우리는 공간적 거리를 정의한다. 이 공간적 거리는 동일한 TOI에서 통과한 위치의 차이로 다음과 같이 정의되어진다.

[정의 5] 궤적들 간의 공간적 거리

도로 네트워크 상에 동일한 시간 간격을 가지는 두 이동 객체 궤적을 TR_A 와 TR_B 라 가정하자. 이때, 두 궤적 TR_A 와 TR_B 간의 공간적 거리는 다음과 같이 정의되어진다.

$$dist_S (TR_A, TR_B) = \frac{1}{n} \sum dist_S(p(TR_A, t_i), p(TR_B, t_i))$$



(그림 5) 공간적 거리의 예

(그림 5)는 서로 다른 두 궤적 TR_A 와 TR_B 간의 공간적 거리의 예를 보여주고 있다. 그림에서, 두 궤적 TR_A 와 TR_B 간의 공간적 거리 $dist_S(TR_A, TR_B)$ 는 다음과 같이 계산되어진다.

$$dist_S(TR_A, TR_B) = (3+4+2.5+3.5)/5 = 3$$

3.3 궤적들 간의 시공간 거리를 기반으로 유사한 궤적들을 검색하는 방법

유사한 궤적 검색에서, 이 방법은 매우 간단하고 흥미있어 보인다. 그러나 이 방법을 적용하기 위해서, 우리는 시공간 거리 측정 기반의 정의가 필요하다. 본 논문에서 우리는 시공간 거리($dist_{ST}$)를 공간적 거리($dist_S$)와 시간적 거리($dist_T$)의 합으로 결정한다. 그러기 위해서는 $dist$ (1미터) = $dist$ (α 초)와 같이 공간적 거리와 시간적 거리 간의 동등성이 먼저 정의되어야 한다. 도로 네트워크 상의 두 이동 객체가 속도(V)를 가지고 움직인다는 점을 착안해 볼 때, 그 이동 객체의 속도를 이용하여 우리는 시간적 거리와 공간적 거리 간의 동등성 문제를 해결할 수 있다. 따라서, 시간적 거리와 공간적 거리 간의 동등성 문제는 다음의 식으로 해결이 가능하다.

$$dist_{ST}(TR_A, TR_B) = d'_S(TR_A, TR_B) + (d_S(TR_A, TR_B),$$

여기서, $d'_S(TR_A, TR_B) = (V_{TRA} - V_{TRB}) \times dist_T(TR_A, TR_B)$ 위 식을 기반으로, 궤적들 간의 시공간 거리는 공간적 거

리와 시간적 거리의 합으로 다음과 같이 정의되어진다.

[정의 6] 궤적들 간의 시공간 거리

도로 네트워크 상의 두 이동 객체 궤적을 TR_A 와 TR_B 라 가정하자. 이때, 두 궤적 TR_A 와 TR_B 간의 시공간적 거리는 다음과 같다.

$$dist_{ST}(TR_A, TR_B) = \alpha \cdot dist_T(TR_A, TR_B) + (1-\alpha) \cdot dist_S(TR_A, TR_B)$$

여기서, α 는 POI기반 거리의 가중치를 의미한다. 이것은 응용 형태에 따라 최우되며, 사용자에 의해 정의되어질 수 있다.

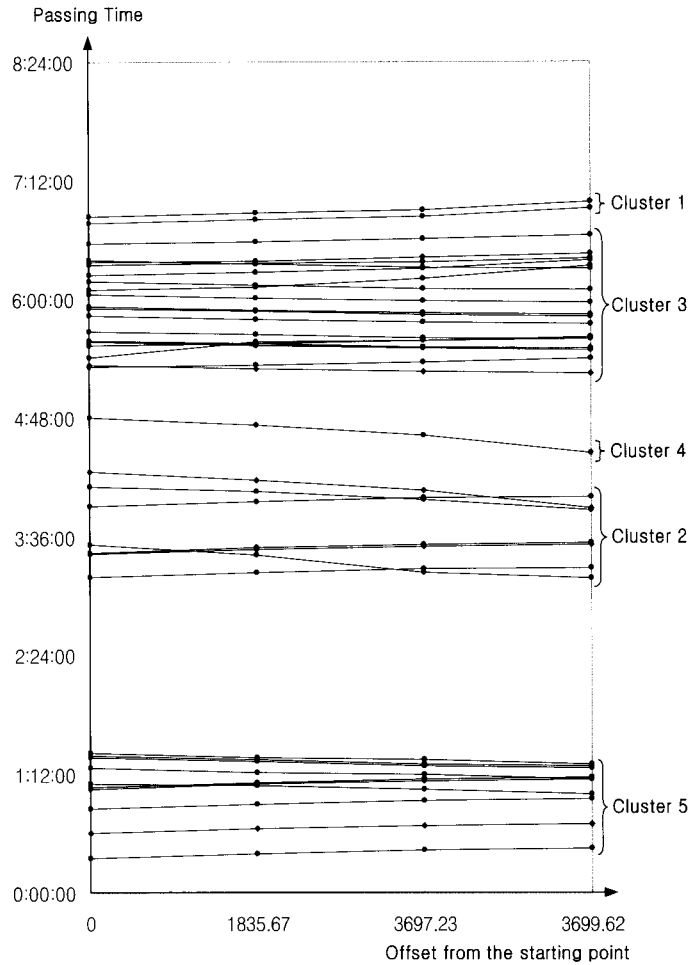
4. 유사한 궤적들에 대한 클러스터링

각 궤적은 다차원 공간에 있는 점으로 표현이 가능하다. 따라서, 우리는 다차원 공간의 점들에 클러스터링 방법을 적용할 수 있다. 우리는 이 다차원 공간을 시간적 궤적 공간이라 부른다. 많은 방법들이 시간적 궤적 공간에 있는 점들을 클러스터링하는 데 제시되어 왔다[13-15]. 우리는 이전의 장에서 제시된 유사한 궤적들을 검색하는 방법 중에서 방법 1을 바탕으로 유사한 궤적들에 대하여 클러스터링을 수행한다.

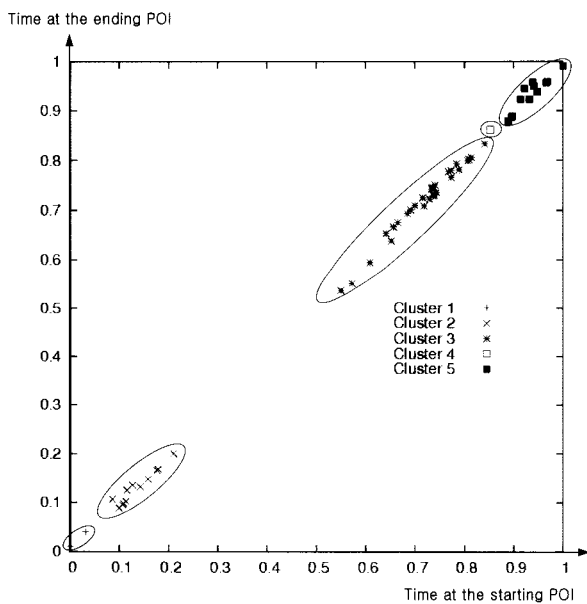
첫번째 단계에 대하여, 우리는 주어진 질의가 포함하는 동일한 POI들을 순서적으로 통과하는 모든 궤적들을 찾는다. 그리고 두번째 단계에서, 우리는 [15]에서 제시된 shifted hilbert curve를 사용하여 시간적 궤적 공간에 있는 점들을 가지고 클러스터링을 수행한다. 앞에서 언급하였듯이 각 궤적은 다차원 공간의 점으로 표현되어지기 때문에 클러스터링이 가능하다. 본 논문에서, 우리는 단순성의 이유로 L_2 거리를 사용하지만 다른 거리의 형태가 사용될 수 있다는 것을 언급한다. (그림 6)은 주어진 질의에 포함되는 POI들을 통과한 궤적들을 검색한 첫번째 단계의 결과를 보여주고 있다. 여기서, x 축은 궤적들의 *offset*을 의미하고 y 축은 그것들의 통과 시간을 의미한다.

(그림 6)에서 아래로 향하는 선은 역방향을 가진 궤적들을 표현한다. 오른쪽에 표시된 클러스터들은 두번째 단계에 의해 획득된 것이다. 즉, 이 클러스터들은 시간적으로 유사한 궤적들을 클러스터링한 결과이다.

우리는 이 실험 결과가 우리의 직관적인 클러스터링과 일치하는 것을 보인다. 그리고 이러한 궤적들은 (그림 7)에서 보듯이 시간적 거리에 의하여 클러스터링되어진다. 여기서, 출발 및 도착 POI들은 단순성의 이유로 x 축과 y 축으로 각각 묘사되어진다. 그리고 x 축과 y 축은 $[0, 1]$ 로 표준화되었다. 흥미있는 요소는 시간적 궤적 공간에 있는 대부분의 점들이 직선 그래프 주위에 발견된다는 것이다. 사실상, 이 선의 기울기는 우리의 실험과 유사한 이동 객체들의 속도를 표현한다.



(그림 6) 첫번째 단계의 클러스터링의 결과



(그림 7) 시간적 거리에 의한 궤적들의 클러스터링

5. 실험 결과

우리 방법의 실현 가능성을 조사하기 위하여, 우리는 부

산시 도로 네트워크로부터 생성한 궤적 데이터 집합을 가지고 실험을 수행하였다. 우리는 도로 네트워크 상에 10,000개의 POI를 정의하였으며, 도로 네트워크 상에서 100,000개의 이동 객체 궤적들을 생성하였다. 그리고, 우리는 추출한 이동 객체들의 궤적 데이터로부터 10,000개의 질의를 생성하였다.

실험에 의하여 유사한 궤적을 검색한 결과들 중에서, <표 1>과 <표 2>는 각각 방법 1과 방법 3에서 검색된 유사한 궤적의 결과를 예를 들어 보여주고 있다. 방법 1과 방법 3은 질의에 대하여 검색된 궤적 결과의 일치율이 상당히 높았다.

그러나 방법 2는 시간적 여과 단계에 의해 너무 많은 후보 궤적들을 가지고 오고 그에 따른 검색된 결과도 너무 많았다. 그 검색된 결과 중에는 의미없는 궤적도 많아 비교 실험에서 제외시켰다.

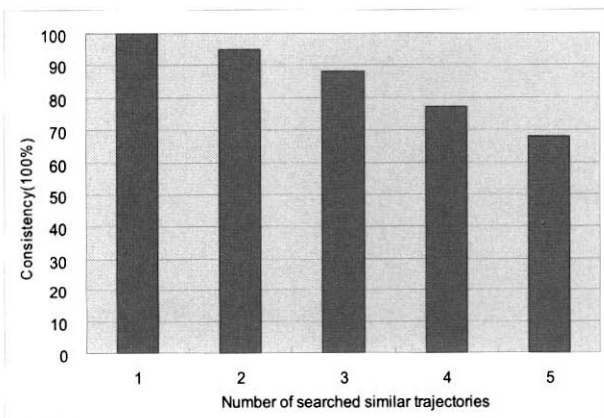
(그림 8)은 위의 결과로 나타난 테이블로부터 두 방법에 대해 검색된 유사한 궤적이 얼마나 정확하게 일치하는지에 비율을 보여주고 있다. 다시 말해서, 이 그림은 각 방법에 의해 검색된 궤적이 일치하는 비율을 보여주고 있다. 그림에서 보듯이, 유사한 궤적 검색에서, 방법 1과 방법 3은 검색된 궤적 결과에 대하여 상당히 비슷한 일치율을 나타내었다.

〈표 1〉 방법 1에 의하여 검색된 유사한 궤적

질의 ID	검색 ID	시간적 거리
...
7	7	0
7	79986	4
7	17298	49
7	35286	49
...
996	996	0
996	18206	4
996	28212	25
996	57414	36
...

〈표 2〉 방법 3에 의하여 검색된 유사한 궤적 결과

질의 ID	검색 ID	공간적 거리	시간적 거리	시공간거리
...
7	7	0	0	0
7	79986	200.676	4	218.935
7	35286	290.2	49	513.872
7	85149	640.774	225	1667.84
...
996	996	0	0	0
996	18206	363.35	4	419.975
996	57414	346.665	36	856.285
996	76413	2914.93	324	7501.51
...



(그림 8) 방법 1과 방법 3간의 검색 궤적에 대한 일치율

6. 결론 및 향후 연구

도로 네트워크 상의 궤적들 간의 유사성 분석은 많은 잠재적인 응용을 가지고 있다. 예를 들면, 이러한 유사성 분석은 차량들의 궤적과 도로 위의 운전자들의 정보 등의 연계를 통하여 타겟 마케팅과 같은 응용 분야에 도움을 줄 수

있다. 본 논문에서, 우리는 도로 네트워크 상의 궤적들의 중요한 특징들과 도로 네트워크 상의 유사한 궤적들을 검색하기 위한 방법들을 제시하였다.

우리의 방법은 두 가지 측면에서 기존의 방법들과 다르다. 첫째, 우리의 방법은 POI와 TOI를 기반으로 도로 네트워크 공간의 특성들을 충분히 활용하는 반면에 기존의 접근들은 유클리디언 공간을 가정으로 하고 있다. 아마도 우리의 방법은 도로 네트워크 공간 상에서 유사한 궤적을 검색하는 첫 방법이다. 둘째, 공간적 및 시간적 유사성들이 우리의 방법에 의해 유사한 궤적을 검색하는 반면에 기존의 방법들은 오직 공간적 유사성만을 고려한다. 그리고 우리의 방법에 의해 검색된 유사한 궤적들은 클러스터링에 적용될 수 있다.

본 연구는 단지 도로 네트워크 상의 궤적들의 유사성 검색을 위한 연구의 시작이다. 우리는 도로 네트워크 상의 이동 객체 궤적들 간의 세밀한 패턴 분석과 클러스터링 분석을 수행할 것이다. 그리고 도로 네트워크 상의 궤적들과 운전자 속성 정보 등을 통합하여 지리 마케팅이나 보험업과 같은 실제 응용에 사용할 것이다.

참 고 문 헌

[1] Dieter Pfoser, Christian S. Jensen, and Yannis Theodoridis. "Novel Approaches in Query Processing for Moving Object Trajectories", In *Proceedings of the 26th International Conference on Very Large Data Bases*, pp.395-406, 2000.

[2] Michail Vlachos, George Kollios, and Dimitrios Gunopulos. "Discovering Similar Multidimensional Trajectories", In *Proceedings of the Eighteenth International Conference on Data Engineering*, pp.673-684, IEEE Computer Society, 2002.

[3] Yutaka Yanagisawa, Jun ichi Akahani, and Tetsuji Satoh. "Shape-Based Similarity Query for Trajectory of Mobile Objects", In *Proceedings of the Fourth International Conference on Mobile Data Management*, pp.63-77, Springer-Verlag, 2003.

[4] Laurynas Speicys, Christian S. Jensen, and Augustas Kligys. "Computational Data Modeling for Network-constrained Moving Objects", In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems*, pp.118-125, 2003.

[5] Michalis Vazirgiannis and Ouri Wolfson. "A Spatio temporal Model and Language for Moving Objects on Road Networks", In *Proceedings of the Seventh International Symposium on Spatial and Temporal Databases*, pp.20-35, Springer-Verlag, 2001.

[6] Ramaswamy Hariharan and Kentaro Toyama. "Project Lachesis: Parsing and Modeling Location Histories", In *Proceedings the third International Conference, GIScience*, pp.106-124, Springer-Verlag, 2004.

[7] K. Hornsby and M. Egenhofer. "Modeling Moving Objects over Multiple Granularities", *Annals of Mathematics and Artificial Intelligence*, 36: 177-194, 2002.

[8] Nico Van de Weghe, Anthony G. Cohn, Peter Bogaert and Philippe De Maeyer. "Representation of Moving Objects along a Road Network", In *Proceedings of the twelfth International Conference on Geoinformatics*, pp.187-194, 2004.

[9] Nirvana. Meratnia, Rolf A. de By. "Representation in Location-based Services: Problems & Solution", In *Proceedings of the Third International Workshop on Web and Wireless Geographical Information Systems*, 2003.

[10] Michail Vlachos, Dimitrios Gunopulos, and George Kollios. Robust "Similarity Measures for Mobile Object Trajectories", In *Proceedings of the Thirteenth International Workshop on Database and Expert Systems Applications*, pp.721-728, IEEE Computer Society, 2002.

[11] Choon-Bo Shim and Jae-Woo Chang. "Similar Sub-Trajectory Retrieval for Moving Objects in Spatio-temporal Databases", In *Proceedings of the Seventh East European Conference on Advances in Databases and Informations Systems*, pp.308-322, Springer-Verlag, 2003.

[12] Lei Chen, M. Tamer Ozsu, Vincent Oria. "Robust and Fast Similarity Search for Moving Object Trajectories", In *Proceedings of the 24th ACM SIGMOD International Conference on Management of Data*, pp.491-502, 2005.

[13] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. "BIRCH: An Efficient Data Clustering Method for Very Large Databases", In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pp.103-114, ACM Press, 1996.

[14] Haixun Wang, Wei Wang, Jiong Yang, and Philip S. Yu. "Clustering by Pattern Similarity in Large Data Sets", In *Proceedings of the 2002 ACM SIGMOD International conference on Management of data*, pp.394-405, ACM Press, 2002.

[15] Swanwa Liao, Mario A. Lopez, and Scott T. Leutenegger. "High Dimensional Similarity Search with Space Filling Curves", In *Proceedings of the seventeenth International Conference on Data Engineering*, pp.615-622, IEEE Computer Society, 2001.

[16] Cyrus Shahabi, Mohammad R. Kolahdouzan, and Mehdi Sharifzadeh. "A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases", In *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, pp.94-100, 2002.

[17] Christian S. Jensen, Jan Kolarvr, Torben Bach Pedersen, and Igor Timko. "Nearest Neighbor Queries in Road Networks", In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems*, pp.1-8, 2003.

[18] Man Lung Yiu, Nikos Mamoulis and Dimitris Papadias. "Aggregate Nearest Neighbor Queries in Road Networks", *IEEE Transactions on Knowledge and Data Engineering*, 17(6), pp.820-833, 2005.

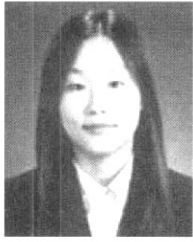
[19] Dimitris Papadias, Jun Zhang. "Query Processing in Spatial Network Databases", In *Proceedings of Very Large DataBases Conference*, pp.802-813, 2003.

[20] 지정희, 최보윤, 김상호, 류근호. "이동객체의 궤적에 대한 연속 최근접 질의 처리", 한국정보과학회 논문지 D 데이터베이스, Vol.31, No.05, pp.0492-0504, 2004. 10.



황 정 래

e-mail : jrhwang@isel.cs.pusan.ac.kr
 1999년 경성대학교 컴퓨터공학과(학사)
 2001년 부산대학교 지형정보학과(석사)
 2003년 부산대학교 지형정보학과 박사수료
 관심분야: GIS, 시공간 데이터마이닝, gCRM



강혜영

e-mail : hykang@isel.cs.pusan.ac.kr
2000년 부산대학교 전자계산학과(학사)
2002년 부산대학교 전자계산학과(석사)
2005년 부산대학교 컴퓨터공학과 박사수료
관심분야: GIS, 시공간 데이터베이스,
Ubiquitous Computing



이기준

e-mail : lik@pnu.edu
1984년 서울대학교 계산통계학과(학사)
1986년 서울대학교 계산통계학과(석사)
1992년 프랑스 국립응용과학원(INSA)
전자계산학과(박사)
1990년~1991년 프랑스 Logicim사
선임엔지니어
1993년~현재 부산대학교 정보컴퓨터공학부 교수
관심분야: 시공간 데이터베이스, 텔레메틱스, 유비쿼터스 컴퓨팅