

# 최대 중첩구간을 이용한 새로운 GPS 궤적 클러스터링

## A New Clustering Algorithm for GPS Trajectories with Maximum Overlap Depth

### 요약

최근 중요한 문제가 되는 GPS 데이터를 이용한 연구는 아주 중요하며, 앞으로 엄청난 효과를 볼 것으로 예상된다. 그중 하나로는 도로생성시스템이다. 도로생성시스템에서 중요한 것은 중심도로를 찾는 것이 가장 중요한 작업이다. 중심도로를 탐색하기 위해선 아주 임의로 퍼져있는 GPS 데이터를 클러스터링 시킨 궤적데이터가 필요하며, 이러한 문제를 해결하기 위하여 본 논문에서는 최대 중첩구간 탐색과 클러스터링 시킨 궤적의 삭제 및 업데이트 과정을 통하여 효과적으로 궤적을 클러스터링 하는 방법을 제안한다. 또한, 실제 대용량의 강남구, 성남시, 서울시 전체를 지나다니는 택시 GPS 데이터를 수집하여 실험을 하였고, 그 결과 빠른 수행시간과 안정성을 보였다.

### 1. 연구동기

최근 차량 GPS 데이터를 쉽게 획득할 수 있음에 따라, 도로생성시스템의 연구가 활발히 이루어지고 있다. 이러한 도로생성시스템의 장점은 지도제작의 비용을 줄일 수 있으며, 제작하는 데에 소요되는 시간 또한 줄일 수 있는 것이 큰 장점이다. 도로생성시스템 연구의 주된 목적은 GPS 데이터를 이용해 자동으로 도로를 생성하는 것이며, 가장 핵심이 되는 작업은 역시 중심도로를 추출하는 것이다. 중심도로(대표궤적)를 추출하기 위한 연구는 활발히 진행되어 왔다[1,2]. 그림 1은 실제 서울시를 지나는 106,787대의 택시 GPS 데이터를 나타낸 것이며, 이를 이용하여 도시 전체의 도로를 효과적으로 생성할 수 있는 것이 최근 중요한 문제가 되는 도로생성시스템 연구이다[3~4].

도로생성시스템에서 가장 중요한 작업인 대표궤적 추출을 수행하기 위해서는 사전작업으로 궤적 클러스터링 작업이 필요하다. 기존연구에서는 클러스터링을 수행하기 위하여 격자단위로 전체 지도를 분할하는 방법[5,6]과 세그먼트별로 나누어 클러스터링을 하는 방법[7]이 있다. 그림 2는 전체 지도를 분할하는 방법[6]을 나타낸 그림이다. 이 방법은 원래의 궤적을 격자로 나누기 때문에 클러스터링작업을 완료하여도, 이후 이어질 대표궤적을 추출한 뒤 다시 대표궤적을 연결해 줘야 하는 비효율적인 작업이 필요한 단점이 있다. 따라서 본 논문에서는 위와 같은 단점을 보완한 새로운 클러스터링 기법을 제안한다. 2장에서는 클러스터링 기법에 대한 세 가지 단계를 설명한다. 3장에서는 실제 강남구, 성남시, 서울시를 지나는 실제 차량 GPS 데이터로 실험한 결과를 살펴본다.

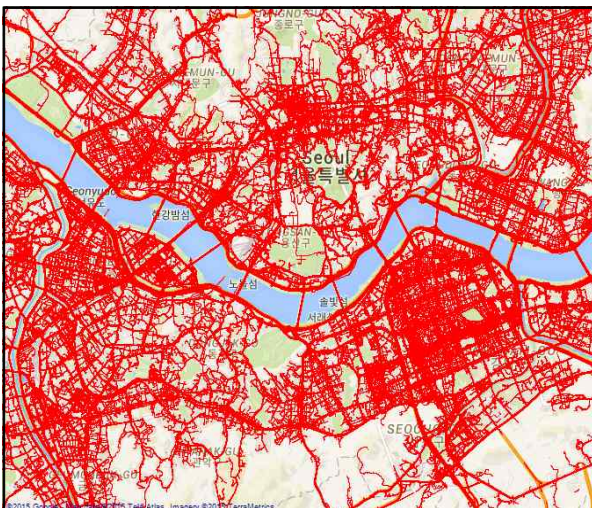


그림 1. 실제 서울시를 지나는 106,787대의 택시 GPS 데이터. 빨간색 선은 GPS 포인트를 연결한 궤적들의 모음이며, 이러한 궤적들은 아주 임의의 형태로 넓게 퍼져 있다.

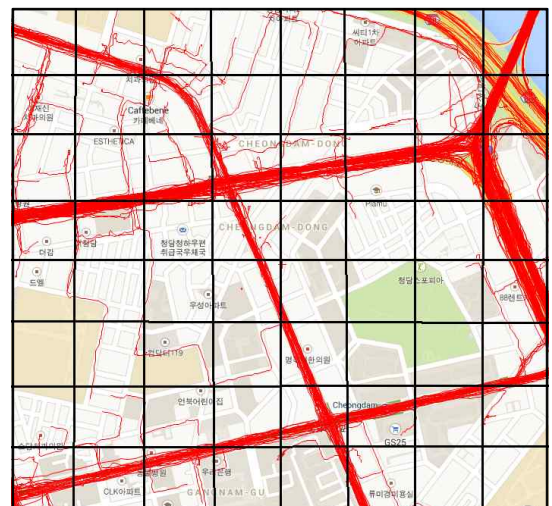


그림 2. 전체지도도를 특정단위로 분할하여 격자마다 클러스터링한 방법이다[6].