

(KCC2015 우수논문)대용량 GPS 궤적 데이터를 위한 효율적인 클러스터링[†]

(An Efficient Clustering Algorithm for Massive GPS Trajectory Data)

요약 도로지도 생성은 인공위성 촬영이나 현장실사를 기반으로 한다. 그리하여 도로지도를 생성하고 수정하는데 많은 시간과 비용이 든다. 이러한 이유로 차량 GPS 데이터를 이용해 도로지도를 생성하는 연구가 활발히 진행되고 있다. 도로지도 생성 연구에서 가장 중요한 문제는 주도로와 같은 대표궤적을 추출하는 것이다. 대표궤적 추출을 수행 할 때에는 시작과 끝이 비슷한 궤적데이터들의 집합을 진제로 하여 궤적을 추출한다. 따라서 대표궤적을 추출하기에 앞서 전 처리과정으로 궤적 클러스터링 작업이 필요하다. 본 논문에서는 이러한 문제를 해결하기 위해 하나의 영역을 일정한 격자로 분할하였고, Sweep Line 알고리즘을 응용해 유사궤적들을 탐색한다. 마지막으로 Frechet 거리를 이용하여 궤적 간 유사도를 계산하였다. 실제로 서울의 강남구 지역에 있는 500대의 차량 GPS 궤적을 가지고 클러스터링 작업을 수행하였다. 또한, 실험을 통해 하여 격자분할 접근방식의 빠른 수행시간과 안정성을 보였다.

키워드 : 궤적 클러스터링, GPS 데이터, 도로 지도 생성, 라인 스위프

Abstract Digital road map generation is mainly based on artificial satellite photographing or in-site manual survey work. So those map generation procedures require a lot of time and budgets to create, update road maps. For these reason people tries to develop an automated map generation systems using GPS trajectory data sets which are obtained by public vehicles. A fundamental problem in this road generation procedure is to extract representative trajectory such as main roads. Extracting the representative trajectory requires the base data set of piecewise line segments(GPS-trajectories) which have near closer starting points and ending points. So geometrically similar trajectories are selected for clustering before extracting one of representative trajectory among them. This paper proposes a new the divide and conquer approach by partitioning the whole map region into a regular grid sub-spaces, then we try to find the similar trajectories by sweeping. Also we applied Frechet distance measure to compute the similarity of pair of trajectories. We conducted experiments using a real GPS data which has more than 500 vehicle trajectories obtained from GangNamKu, Seoul. Experiment showed that our grid partitioning approach work fast and stable to be used in real application for vehicle trajectory clustering.

Key words : Trajectory clustering, GPS data, Road map generation, Line sweeping

1. 서론

최근 GPS수신 단말기 보급(스마트 폰, 네비게이션)의 증가에 따라 GPS 데이터를 쉽게 얻을 수 있고, 이에 따라 GPS 데이터를 이용한 연구가 활발히 진행되고 있다. 관련연구에서 [1]은 궤적 클러스터링 방법을 제시하였고, [2]는 GPS 데이터를 이용한 연구들의 성능을 비교 평가하였다. 이러한 연구들 중 하나는 도로 생성 시스템으로 이 시스템은 사용자로부터 받은 GPS 데이터를 이용하여 해당 도로를 자동으로 찾아주는 시스템이다. [3]에서는 궤적들을 Clarify하여 Routing Map을 제시하였다. [4]역시 마찬가지로 GPS 데이터를 이용해 실시간으로 도로를 생성해주는 시스템을 제안하였다.



그림 1 강남구 지역에서 수집된 택시 GPS 데이터. 빨간색 선은 총 3000대의 택시의 궤적을 나타내고 있다.

Fig. 1 Taxi GPS data collected from GangNamKu. The red line represents trajectories of 3000 taxis.

그림 1은 강남지역을 대상으로 택시 GPS 데이터를 받