Project Summary

Increasingly, location-aware datasets are of a size, variety, and update rate that exceeds the capability of spatial computing technologies. This project addresses the emerging challenges posed by such datasets, which we call Spatial Big Data (SBD). SBD examples include trajectories of cell-phones and GPS devices, vehicle engine measurements, temporally detailed (TD) road maps, etc. SBD has the potential to transform society. A recent McKinsey Global Institute report estimates that personal location data could save consumers hundreds of billions of dollars annually by 2020 by helping vehicles avoid congestion via next-generation routing services such as eco-routing. Eco-routing may leverage various forms of Spatial Big Data to compare routes by fuel consumption or greenhouse gas (GHG) emissions rather than total distance or travel-time.

However, the envisaged SBD-based next-generation routing services pose several challenges for current routing techniques. First, SBD requires a change in frame of reference, moving from a global snapshot perspective to the perspective of an individual object traveling through a transportation network. Second, SBD magnifies the impact of partial information and ambiguity of traditional routing queries specified by a start location and an end location. For example, traditional routing identifies a unique (or a small set of) route(s), given historical and current travel-times. In contrast, SBD may identify a much larger set of solutions, e.g., one route each for thousands of possible start-times in a week, significantly increasing computational costs. Third, SBD challenges the assumption that a single algorithm utilizing a specific dataset is appropriate for all situations. The tremendous diversity of SBD sources substantially increases the diversity of solution methods. For example, methods for determining fuel efficient routes leveraging engine measurement and GPS track datasets may be quite different from algorithms used to identify minimal travel-time routes exploiting temporally detailed roadmaps. Newer algorithms may emerge as new SBD becomes available, creating the need for a flexible architecture to rapidly integrate new datasets and associated algorithms.

Intellectual Merit: This project is expected to result in III innovations in three areas. First, Lagrangian frame of reference, a novel concept in computer science, will be explored at conceptual, logical and physical database levels to model travelers’ frame of reference, a major departure from traditional snapshot-based approaches. Second, to address increased computational cost from partial query specification, we will explore the concept of route-collections, and scalable algorithms for finding route-collections. For example, to identify a route-collection over all possible start-times of a given time-interval, we will investigate a critical time point approach which divides a given time-interval into a set of disjoint sub-intervals of stationary-rankings among alternative routes. The approach is not only novel but also very important for the field. We believe that critical time points may become a vital component of dynamic programming (DP) solutions, which will need reconsideration in the face of emerging SBD that violate DP assumptions about stationary ranking of alternative solutions. Third, to address the increasing diversity of SBD methods, we will investigate algorithm-ensembles, flexible architectures that allow rapid integration of new data sources and routing algorithms. The team has a track record of publications and innovation not only on the current generation of routing algorithms and digital roadmaps, but also on emerging TD roadmaps and spatio-temporal routing algorithms. The team also has access to required resources such as TD roadmaps, GPS trace data and fuel consumption datasets.

Broader Impact: The proposed work, if successful, will serve U.S. goals for energy independence and sustainability by laying the ground work for eco-routing and other travel-related services that reduce fuel consumption and greenhouse gas emissions. By increasing the availability of SBD and related software prototypes, the project also enhances the research infrastructure for other researchers, few of which have access to these datasets currently. Educational activities will include curriculum development and training of students in the emerging area of SBD and computational aspects of next generation routing services. The research team includes female and minority graduate students and the PI has a track record of participation in summer institutes involving undergraduate students from historically black colleges and universities. Results will be submitted for publication and presentation at relevant peer-reviewed computer science conferences and journals.

Keywords: Spatial Big Data; Spatial Databases; Spatial Data Mining; Routing; Road-maps